

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2005, 02:48:50 ; Search time 5973 Seconds

(without alignments)  
11292.426 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392  
Sequence: 1 atgaataaccatcatgatac.....ccggcggaagtcgtcgat 1392

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:.\*  
1: gb\_ba:.\*  
2: gb\_ptg:.\*  
3: gb\_in:.\*  
4: gb\_cm:.\*  
5: gb\_ov:.\*  
6: gb\_pat:.\*  
7: gb\_ph:.\*  
8: gb\_pl:.\*  
9: gb\_pr:.\*  
10: gb\_ro:.\*  
11: gb\_sbs:.\*  
12: gb\_sy:.\*  
13: gb\_un:.\*  
14: gb\_vi:.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	628.8	45.2	11493	1	AF283498	AF283498 Azospirillum
C 2	555.6	39.9	300883	1	AE016809	AE016809 Vibrio vulnificans
C 3	541.2	38.9	248850	1	AP005348	AP005348 Vibrio vulnificans
C 4	516.2	37.1	302100	1	AP005087	AP005087 Vibrio vulnificans
C 5	431	31.0	311050	1	BX294133	BX294133 Pteridinium
C 6	171.2	12.3	110000	1	AE017282	AE017282_06 Continuation (7 of
C 7	133	9.6	10676	1	AE008293	AE008293 Agrobacterium
C 8	133	9.6	10707	1	AE009321	AE009321 Agrobacterium
C 9	122.8	8.8	348077	1	AP003000	AP003000 Mesorhizobium
C 10	119.8	8.6	300150	1	AP005945	AP005945 Bradyrhizobium
C 11	118.2	8.5	110000	1	RME591985_07	RME591985_07 Continuation (8 of
C 12	116	8.3	196050	1	AL646058	AL646058 Ralstonia
C 13	113.2	8.1	298700	1	AP005951	AP005951 Bradyrhizobium
C 14	109.6	7.9	12129	1	AE007223	AE007223 Sinorhizobium
C 15	104.8	7.5	299750	1	AP006575	AP006575 Gloeobacterium
C 16	104.4	7.5	339650	1	AP003583	AP003583 Nostoc sp
C 17	100.6	7.2	348068	1	BX572604	BX572604 Rhodospirillum
C 18	98.4	7.1	300700	1	AP006573	AP006573 Gloeobacterium
C 19	98.2	7.1	110000	1	BX950851_33	BX950851_33 Continuation (34 of

20	95.2	6.8	17155	1	AY273169	AY273169 Rhodospirillum rubrum
21	95	6.8	4148	1	KPNASERC	K2741 Klebsiella pneumoniae
C 22	93.4	6.7	13551	1	D90306	D90306 Synecococcus
C 23	89.8	6.5	3253	1	SYONIRNRT	D12723 Synecococcus
C 24	89.2	6.4	300363	1	AE016781	AE016781 Pseudomonas
C 25	87	6.2	10798	1	AE004604	AE004604 Pseudomonas
C 26	86	6.2	300300	1	AP005373	AP005373 Rhodospirillum rubrum
27	84.6	6.1	349142	1	BX572599	BX572599 Rhodospirillum rubrum
28	82.4	5.9	1371	6	AR36463	AR36463 Sequence
C 29	82.4	5.9	299650	1	AP005955	AP005955 Bradyrhizobium
C 30	80.2	5.8	4080	1	SYOCMP	D26358 Synecococcus
C 31	79.4	5.7	298750	1	AP005375	AP005375 Thermosyn
C 32	77.8	5.6	302550	1	AP006581	AP006581 Gloeobacterium
C 33	75.2	5.4	11318	1	AE008005	AE008005 Agrobacterium
C 34	75.2	5.4	11330	1	AE009039	AE009039 Agrobacterium
C 35	74.6	5.4	6449	1	PLNRTABC	Z1958 P. lammosum
C 36	74.4	5.3	347400	1	AP003591	AP003591 Nostoc sp
C 37	73.8	5.3	7520	1	AE005736	AE005736 Caulobacter
C 38	72	5.2	132106	1	SYCSLTH	D64006 Pseudomonas
C 39	71.8	5.2	310325	1	AE016864	AE016864 Pseudomonas
C 40	70	5.0	3218	1	AF001333	AF001333 Synecococcus
C 41	69.2	5.0	11321	1	AE005737	AE005737 Caulobacter
C 42	65.8	4.7	3719	1	AY283674	AY283674 Cyanobacter
C 43	63.4	4.6	2210	1	AVNASST	X83602 A. vineandii
C 44	63.4	4.6	4080	1	SPCCNRT	X61625 Synecococcus
C 45	60.8	4.4	110000	1	CP000011_11	Continuation (12 of

## ALIGNMENTS

RESULT 1	AF283498	11493 bp	DNA	linear	BCT 02-JUL-2002
LOCUS	Azospirillum brasiliense	nitrate assimilation gene cluster, complete sequence.			
DEFINITION	AF283498				
ACCESSION	AF283498.1	GI:21666012			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene					
CDS					
FEATURES					
source					
gene				</	

AVDLPRPRDRTAVNHDPDFKRLAEVTAMLEAGASRAAAAVDESLTMAPATPTAP  
PKAYLAGOGPGGLSDRYVEFSRVTKTPATPKPGLTVNDGFDLKKRKEGFIILIGSG  
CGKSTVLNVAAGLADVSGSIVLDGKEVVGAGDPRVVPQAPSLPFWLALNVAWVGV  
DRVYPOAGGERNDIARYILSRVGLNMAHKTAELSNKQKRVGIARAPALSPKILL  
LDEPFWMLDSLTKMELQEVLEWVWARTQVATLCVTHDVEAILLADRVMVMSNGPNAR  
IGHILLEVDI.PHPRTQALDEHPRYAVREBLNLPFGGHHGVKAA"

gene  
/gene="naab"  
/complement(2204. .3295)  
/codon\_start=1  
/gene="naab"  
/protein\_id="AA073542.1"  
/product="nitrate transporter integral membrane protein"  
/db\_xref="GI:2166014"

CDS  
/translation="MTSVTBASTLSAEMRARRKRLAHLNRTAATQIYGLGMLS  
PLIRMAAGSPROQGOELMOQKVPPLTALMLPLAAMVLAPOVNTSLGAIIPBAQWE  
OIGILHADHKAERAKEAAFYDRQAKNABLPLADPAEAVKTRPVAKGPIQIILSL  
KIVFTGLGALVAAPLVGACLSRTVNAAMNPLQIPKVPSPALPLVTVMVSTV  
SSDNPFEEKSLTSAVTTLCSLMPILNTAVGVSSIDKDLNMGKVLQSGPTWVR  
LVLSALPYIFGLRLSLGVMVLLAEMLAQNPLGKFWDEPQNGSSSLARIMV  
AVFTGLIGFLDRVWLAFOAVSHGTG"

gene  
/gene="naab"  
/complement(3455. .4843)  
/codon\_start=1  
/gene="naab"  
/protein\_id="AA073543.1"  
/product="nitrate transporter substrate-binding protein"  
/db\_xref="GI:2166015"

CDS  
/translation="MIRLPASPTALLSNAATLMLGSAQAAPLDEKQDLGFIK  
LTDMAFLAIAEKGFEDGLSVTLERQANMKNLPLKSGDGLHMKDALGPIQIIL  
GGTQANVVTAPSMIDNGITLSNEVMKMDNLPKSGDGLHMKDALGPIQIIL  
YRABEGRPTMGKVPVSTHNYELRYTLAAGINPYYAENDVSGQIQADALISVTPP  
QMPLEAGTIFGYSVGEPMNQAVKGIQVPIYDTEIKNNPEKVFSGTGMMAKN  
PKTHLALVRLIRAAWMDENGANNAEAVKILAKSEYVGAQKVIANSMTGTFEYK  
GKRAPVDPNVFERNATYVPYSDAVWVLTQNRPMQOIAEAKPDAYDSTARKVKE  
IYKAARLLVEBGEAKADPMTSDGYKFLDNGFIDGIAYDGRKPNVYTLKPIGLKG  
GGAUVGQGLVGC"

gene  
/gene="naab"  
/complement(5514. .7979)  
/codon\_start=1  
/gene="naab"  
/protein\_id="AA073544.1"  
/product="assimilatory nitrite reductase large subunit"  
/db\_xref="GI:2166016"

CDS  
/translation="MSTHNPPTPRERLVVVGNGMAGIRTLSELAKADRYDITVFG  
AEPHPVNRIMLSPLVLAGETPEOIVLNGRDWYEANGIRLITGDRREVLDKANCVTYA  
ISGLTVPYKLLIATGSTPVII PVPGSTLGVGVPDIALVDVPMLEAARGRAVIG  
GGILGEBANGKLKVMQMDYTVLHMDTLERQDLRSAGALLHHELEKRGITVLTGATD  
AELVSEKVSARLKNQDELPAULVMAVGIRPNTTLGAAGLACERGIQVDMATTS  
DPAIILSVGECEBHRGQTGIVAPLFEMAVVAARLAGSEAAVLTGSVTSTKLKVTG  
VSAAGDFTGKDCEDIVFRDAARGVYKRVVVENRILGALVGDTCGDMYQOMKQD  
TEVAPVBDTLIFGOSFGGAANPAAVAPLPSAELICGNGVCKGTITATLBERKLS  
SLDEVAHHTKASAGSCSGOYVOLLALTLPDGYAEGAKAKMCKCTDPTDTHVRAT  
TALDLKTPDVNQRLEKRTPDGCHNRPALNTYLCEMFGETKIDRSRYTNERVHAN  
IQKDGTVSVPRMGGTISAKELRATADVDFKALPTVATGQQRIDLEGVKEDLPI  
WADLNAAGVMSGHAYAKGLRTVCTGVSEBMRFGTDGTGLKERTMTDTPHPK  
VTLAVSGCPRNCEATIKDLGVCSVDSELEHVGAGNVAACDLVLEVEAELE  
YTGAVVQLYREEARLYERTAPWYERGVGLHIRRLVDADEGSAALANARPLFSQPSQD  
DPAERANGSVDRHEFAPLAQYG"

gene  
/gene="naab"  
/complement(7986. .8342)  
/codon\_start=1  
/gene="naab"  
/protein\_id="AA073545.1"  
/product="assimilatory nitrite reductase small subunit"  
/db\_xref="GI:2166017"

gene  
/gene="naab"  
/complement(8473. .11130)  
/codon\_start=1  
/gene="naab"  
/protein\_id="AA073546.1"  
/product="assimilatory nitrate reductase catalytic subunit"  
/db\_xref="GI:2166018"

CDS  
/translation="METRTTCPCYGVGGLVASYSDGTGVAGDPDPHARNGLCS  
KTNLDPRLGLAEIRLHPQVGRVYAMPEALAEVVKRSLDPTAKGPPSVARVSGOL  
LITEDYAVNKLAKGFLGTANTIDMSRLCMASVAGKKGFGADVAPVYDDIDDEADLV  
LLVGSNLMACHPILNQRLIAAKHRTGTLVLDPEPRTYAVDADNHLPRPCTDLALF  
NGLIVYLDNHLADNGLIDARFVAHETGLAEALAEARADAPSPAHVAACGLPLAEV  
EAFYAEVTVRRTTVVYSGVNOSSQGTDTVNAILNHFILNRIAPGMPSPVSTGQP  
NAMGREGVGLANOLAAHMGFEPEVDELRFRPMNAPRYADKGLKAVDLFRAVEGTV  
RAVMVATNPVMSMDARVRBALAKCTVIVSDCVADPTDPAIRSHIRLPAWSEKD  
GTVNDSMTTISRORAFRLPAGBARPDWIVTVQVAMGPRPAAAFSETPADVREHALL  
SAFENGGRAPDITGGLAGLSDLEFETTLPAEPRPAGCPTRRLPFDGCFEYDEGRAR  
FVPAVSRLPLPSVSEPSLMLNTGRLRDQMTTRTGLSPRLSAHABEPLDLHPQDA  
ARGLTDGALVSLSTVGSTALARLRTEAQRVEALPMMWTRFTASCVYGLVDLA  
VDPVSGOPDLKRMPTVTLTPYPAAMTGFLLAEPEVPAAGVYRSRAVPGHLELASE  
DALPAPETLARGFAGATVYDGAARGVLRAMVHTDRLBACILFSLSGALPBRGLVE  
LLSARLDGAARRALDSRSPVPTADBERTYWSCGVGVNRLIAIQSOGLLTVVQEVG  
AALRGVTCGSLPELKEIILADQKSHVA"

ORIGIN

Query Match	45.2%; Score 628.8; DB 1; Length 11493;
Best local Similarity	68.7%; Pred. No. 8.5e-111;
Matches	912; Conservative 0; Mismatches 407; Indels 9; Gaps 3;
QY 69	GGGTTTGAAGATTTGGCCCGATGTCGGG3CAGATTG3CAAGCTGAAAGAAAGATTGAA 128
DB 4787	GGCGTTGATGTCGGGTCGCGCCAGCGCTCCGCTCGACGAGAAAGAACAGCTCA 4728
QY 128	ATTGCGCTTCAATCAAGCTCACCGACATGCGCGCGTGGCGCGCGCGCGGAAAAGCTT 188
DB 4727	GCTGGCTTCAATCAAGCTGACGATATGGCCCGCGCTCGCATGCGCGGAAAGGCTT 4668
QY 188	CTTGAAGACGAGGCGCTGTTCTGTCGAACTGGAACGCAAGCGCAACTGGAAGGTGAT 248
DB 4667	CTTGAAGACGAGGCGCTGTTCTGTCGTAACCTGGAAGCGCAAGCTGGAAGGTGAT 4608
QY 249	GGATAGGGTCTGAATGCGGAACTGGAACGCTGCGACATGCTGGCGCGCGCGCTTAC 308
DB 4607	CGACGCTGATCTCGGGGAGCTGGAAGCGCGCCACATGCTGGCGCGCGCGCTGAG 4548
QY 308	GGCCAGCGTGGCTTCCGCAACCAAGCCGATATGAGGTGCGCTTCAAGATGGCTTCA 368
DB 4547	GGCCAGCGTGGCTTCCGCAACCAAGCCGATATGAGGTGCGCTTCAAGATGGCTTCA 4488
QY 368	CGCAAGCGATTAAGGTGTCGAATGAATCTGGCATGAGATGAAGCGCAACATGAC-- 425
DB 4487	CGCAAGCGATTAAGGTGTCGAATGAATCTGGCATGAGATGAAGCGCAACATGAC-- 4428
QY 426	GCTGAAGCGGTAAACCGGTGATCGATCAAGGCAATTAATCTAAACCGGTGCTGA 485
DB 4427	GGGCGCGGACGCGCAAGCGCTGACCCGATCAAGGCGGACGCGGATCATGCGC 4368
QY 486	AAATPAAAGCGGAAAGCGGCGGCTTGAATATGCGATGACCTTCCGCGCGGATTC 545
DB 4367	CGAGTACCGGAGAGGCGGCGGCGCTTGAATATGCGATGACCTTCCGCGCGGATTC 4308
QY 546	CAACATCAAACTGCTTACTGCTGCGCGCTGCGGCTTCAATCTGCTACTATTCGCG 605
DB 4307	CAATTAAGAGCTGCGCTTACTGCTGCGCGCGCTGCGGCTTCAATCTGCTACTATTC 4251
QY 606	GCGGAGGACATTTTCGCGCAAAATCGGCGGACGCGCTTGTGCTGCTGACCCCGCGCG 665

Db 4250 GCCCAAGAGTGTGCGGCGAGATCCAGGCCGACGCGTCTCCGTCAACCCGCGCC 4191  
 Qy 666 GCAATGCGGTCCAGCGTGGAAAGCCCGACCATTTTCGTGTTGCGTGGCGACCGCG 725  
 Db 4190 GCAAGATGCGGCGACCTGGAGCGCGAGCATCTTCGTGTAAGGCGTGGCGACGCG 4131  
 Qy 726 GAACGACGAGCGCGTTCGAAAGGCGTAGGCGCGTGAATCAGCGATGAAGACTCTG 785  
 Db 4130 GAACGACGAGCGCGTGAATGAAGGCGATCGCGTGGCGTATCAACCGACGAGTCTG 4071  
 Qy 786 GAAGGACGACCGCGGAAAAAGTCTTCGCGGTGACCAAAACAAATGGCGGAGAAATACCCCA 845  
 Db 4070 GAAGAAACAAACCGGAGAGGTTCTTCGCGGTGACGAGGCGTGGCGGCAAGAACCCCA 4011  
 Qy 846 CACCTATCTGGCGGTGACCAAGGCGGTGATTCGCGCGCGATTCGCTGAGCGCCGACAA 905  
 Db 4010 GACCCATCTGGCGGTGACCAAGGCGGTGATTCGCGCGCGATTCGCTGAGCGAGAACG 3951  
 Qy 906 TAAACAAGAACCGCAAGAGGCAATCGAAATGCTGGCGCAAAAACAATACGTCGTCGTA 965  
 Db 3950 CAACGCGCAACCGCGCGGCGGAGCGGTGAAGATCTTCGCGCAAGTCCGAGTATGCGGCGCA 3891  
 Qy 966 CGTGAAGTGTGCGCGGTGACGAGCACTTCGAAATGCAAAAAAGACGATTAAGC 1025  
 Db 3890 CGCCAAAGTGTATCGCCCACTCGATGACCGGAGCTTCGATGAGAGGCGGACAAAGG 3831  
 Qy 1026 CGCGCTACCGGACTTAAACACTTCTTTGGCCAGCGCGCGATTCGCTCTTACAGCAG 1085  
 Db 3830 GCGCGTTCGCGACTTCAACGCTCTTCGCTGACCAAGCGCACCTAACCTTCTTACTCCGA 3771  
 Qy 1086 TSCAGTCTGTATCTGACCCAGCTGAGGCGTGGGCGATGATCAATGAATTAACACCGGA 1145  
 Db 3770 CGCGCTGTGTACTGACCGAGATGCGCGCGTGGGCGAGTTCGCGGACCAAGCGCA 3711  
 Qy 1146 CAATGCTATCTGATACCGCAAGAACTCTTACCGCGGACATCTATCTCGCGCGCGC 1205  
 Db 3710 CGCTGTGTATGACGAGACCGCGCGCAAGGCTTCAAGCGGAGATCTTGAAGGCGCGC 3651  
 Qy 1206 CAAAGAACTGTGTGCAAGAGGCAAGGCGCAAGGCGCAAGCTTCC--TGCCGATACCTC 1262  
 Db 3650 GCGGCTGTGTGTGAGGAGGAGGCAAGGCGCAAGGCGCACTTCCCTGAGCAGGACG 3591  
 Qy 1263 GATCAAGCGCGTGCAGAAATTTCTTCATGCAAAAGTGCCTTCGATGCAACAGCCCA 1322  
 Db 3590 CTACAAAGCGCTGTGATACGATTCATGACGCGCATCGCTACGAGCGCGCAAGCCCA 3531  
 Qy 1323 CGATTACTCTGCGCAAGTTTGCATAGTCTGTAAGGCAAGAACCGTATGCGCGCGCA 1382  
 Db 3530 CGAGTACCTGACGAAGAGCTGCCATCGGCTGAAAGGCGGCGAGCGGTGAGGCA 3471  
 Qy 1383 GGTCTGTG 1390  
 Db 3470 ACTGCTCG 3463

RESULT 2  
 AE016809/c 300883 bp DNA linear BCT 29-SEP-2003  
 LOCUS AE016809  
 DEFINITION Vibrio vulnificus CMCP6 chromosome II section 2 of 6 of the complete sequence.  
 ACCESSION AE016809 AE016796  
 VERSION AE016809.1 GI:27358288  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Vibrio vulnificus CMCP6  
 Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrrio  
 REFERENCE  
 AUTHORS  
 Kim, Y.R., Lee, S.E., Kim, C.M., Kim, S.Y., Shin, E.K., Shin, D.H., Chung, S.S., Choy, H.E., Proguile-Fox, A., Hillman, J.D., Handfield, M. and Rhee, J.H.  
 TITLE  
 Characterization and Pathogenic Significance of Vibrio vulnificus Antigens Preferentially Expressed in Septicemic Patients

JOURNAL  
 PUMED  
 14500463  
 2 (bases 1 to 300883)  
 REFERENCES  
 Jeong, H., Moon, Y.H. and Kim, J.J.  
 Direct Submission  
 JOURNAL  
 Submitted (13-DEC-2002) Genotech Corp., 461-6, Jeonmin-dong, Yuseong-gu, Daejeon 305-811, South Korea  
 3 (bases 1 to 300883)  
 REFERENCES  
 Rhee, J.H., Kim, S.Y., Chung, S.S., Lee, S.E. and Choy, H.E.  
 Direct Submission  
 Submitted (13-DEC-2002) Department of Microbiology, Genome Research Center for Enteropathogenic Bacteria, Chonnam National University Medical School, Hak-1-Dong, Dong-gu, Kwang-Ju 501-746, South Korea  
 FEATURES  
 source  
 1..300883  
 /organism="Vibrio vulnificus CMCP6"  
 /mol\_type="genomic DNA"  
 /strain="CMCP6"  
 /db\_xref="taxon:216895"  
 /chromosome="II"  
 269..577  
 /locus\_tag="VV20274"  
 269..577  
 /locus\_tag="VV20274"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Unknown"  
 /protein\_id="AA007239.1"  
 /db\_xref="GI:27358289"  
 /translation="MILENKPTNSROPILSNLKKODFIYHEKTRFOILKAQNTSTV TISGKDKLKGIVVELITIGSRAAIGICGRALQKDFVDAISLRIDVKEARECVR "  
 565..2553  
 /locus\_tag="VV20275"  
 565..2553  
 /locus\_tag="VV20275"  
 /note="COG2199"  
 /codon\_start=1  
 /transl\_table=11  
 /product="F0G: GQDF domain"  
 /protein\_id="AA007240.1"  
 /db\_xref="GI:27358291"  
 /translation="MCSLSKSLSTIRKLLLPILFETLFLIAQGFHMASTYKDDOL NHRSTLAVGAQNTQASVLFENDQVQWELISFISIDSILVLEKSLBGLVAYRQ REMPVNDHNVLSLPYFETFTYSQDIVEQNALTLSTQTLBDMFTLBSLV VFLATILSGSTLYLIVQKVIDPYKINSGMODLVEGRIKGSLVSHSDMGELVN AFNAMSQRLARSDQVHTLDRLEBEKAPATDVIAVOTALLVNEHGDILHNNAA RMLFOSDAQRIQNHKLSEMLCKEBSFVNALTSKNPLREPLVEGCLPTGSKLLITST RILSKRRATLFSIQDYTEIETALNRQRIAAAGVFENSQGLVLDHNGILITNTNPATR MLGYPOETLISKLKAPDWQOIQTIMPITIGESLEHYQWQGVLEKHYNGNLIEMFA KVSRLSKRVQVEFVDNVIILSDLSGVEMERLEYLAHHDALTGLANRAKLRLEDE LKLCNFTSEFALYMDLDGKRFINDTGADGHDVLKVSERLNEVRSGLVVARLS GDEFVLLIKRADNSVTLLSRILLESLSQPIEYLNQTMQVCSIGVKLVGPEKXMER ILKSADTMYKRAKXSGKOATLISGEVEACLSVTSR"  
 complement(2838..4703)  
 /locus\_tag="VV20276"  
 complement(2838..4703)  
 /locus\_tag="VV20276"  
 /note="similar to mostly Fe transport; COG1629"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Outer membrane receptor protein"  
 /protein\_id="AA007241.1"  
 /db\_xref="GI:27358291"  
 /translation="MEGRKRVKLVGASQPGVPOGPGYSPINSSAVSVDPMLKSVBI VKGASLSHGSDAIGGVVAFPTKQPADLTKGKYVGQAKLSYSEDSKSPFHVAK RFDNVEALVATYTRDQKELQNPSPAPDDIYVESODIYKNDLITLQSLSDHRLER LGEVLYNQNDSDIASSYKFNNSEDITQKQNLGLKHLVLDVPTITDSITSLRTWDKE ENGLITNREKASAGVPPKPNQNGQKQDQVANKLELTQEPKLESTHLYVGLS YKTSRISNTNREPNFSDPNTPKIYVTTDAKEMSGFLPQDEISVANDKLVLVTPGVRY DFTSDPEDVDESIFKPSDAFPGALCATKVAQEGTLFTQISGFPAPSNELYYT YDNGHGVTNNPNPLSEKSLSYEIGRHNHTLSTSLIALYSPDYPFICOTSRG GLTETTYNISEATITKGIENVNTLDNRINLPLQMTTRFPAAYTVGDEGQNPLNSV

NPMNAVGAVNDVADPTLTGCTSIKUNVYTKKNSHNLIDGNNGGTANOVHELPSAIVDID  
 TAYEPETDVTYPRACIINMLTNEKYLIMNDIGEKRLSDLTQARERNITISVKTEF"  
 gene .6094  
 /locus\_tag="VW20277"  
 5180..6094  
 /locus\_tag="VW20277"  
 /locus\_tag="VW20278"  
 /note="COG0583"  
 /codon\_start=1  
 /transl\_table=11  
 /product="transcriptional regulator"  
 /protein\_id="AAO07242.1"  
 /db\_xref="GI:27358292"  
 /translation="MILITKGVNOSPITLTAHLTIDIPRGSFASANLEDAPESS  
 LSYQIKBOBDLMIIPRSGKANFTAGCLIERGTTILAEKLYNDATLILANC  
 ELDISIADVGIIIGHFSPILDALGNVSRTPVRLQELTILAGCSFSLQSGRADLVCKP  
 LEVLDHADVADRTIGMSMTWVAAPNHYVHKRSGEENIDARKEKYIILAIDARQOPAI  
 TVNILOKQPRLVNTVLOAKVDVALVTGLIGTLPDTIATPLIEAGDKIEBAGQCHLE  
 IILAKRNQMGCAKWCIOYKXMKKL"  
 gene comp1ement (6143..8821)  
 /locus\_tag="VW20278"  
 complement (6143..8821)  
 /locus\_tag="VW20278"  
 /note="other COG candidate exists; similar to  
 acetyltransferase; COG0454"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Histone acetyltransferase HPA2"  
 /protein\_id="AAO07243.1"  
 /db\_xref="GI:27358293"  
 /translation="MNRKOLFKPRTSIVIGASORHLNAGNIWKNLISGSPDITIP  
 VTPKRSAGVLAIVSVSTVELPLVPDIAILCTRSEKNVALFKQLAEITGEPIYLAISTT  
 DYPHDDDSVADACLRILARESNNRVLGNSTGLIPQKFNASPVTAKGKIAFTS  
 OSAACITILPWANDKEIGFSAFVSIINGLIDPELDLDYSTDSTKDAIILVYDSITV  
 DARRISADARSRRLILVXKGRSILGRBALNSGADTDIYVDSAIRSGMLRYNVA  
 NHEHFAVETITESTVPRGERLIATITWAGPALMAVDTLPERGQOLANISSEITML  
 SGTSSSHSNHPIDIVGDADKKRYATINMILDSQCADATLIMHSBPAVSDEPTQ  
 AIVDAIQSPKPKRFNILLNWSGEOTAKPARKITLKAGIPYKIPESAVVAFFHLYE  
 RNNQQLWETPTTAAPELHVSDNNAKAVSHLMDKEITQDLDTQLGSLKLFHFDVLA  
 PMVQISLSEAVHIAEQIGPYAVLGRSPDLAHSKDVOGVMLNLSNGEVSASDAIILI  
 DTKLAYSANLHGLVOGMAKLAGSEIRIVYITDKGVFVILIGQGSFMSBARDA  
 TALPLIMTATLVQIVRAIKERGHRLCKLPEPMDVGLAKEIVRSIMVTEIPIKE  
 LDIHVVLANGEFTLIDADLTLENLYGSGQERLIRPPPAFVETMQKDGEPILILRP  
 ILPEBESHAETINNVSKEDLYKRFESFVGSEHGMALNLTQIYIDREMAVVASFESS  
 GSAKIGARALINPENTEAEFALIRSDLKGKGLGKALNKKIIDYCRQKTKTKGTGCM  
 TPTNRGMLTIAQKGFSLDIQFEDGVADNVLPLNP"  
 gene 9046..9621  
 /locus\_tag="VW20279"  
 9046..9621  
 /locus\_tag="VW20279"  
 /note="COG3266"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Uncharacterized protein conserved in bacteria"  
 /protein\_id="AAO07244.1"  
 /db\_xref="GI:27358294"  
 /translation="MKKVAIITLVLSACASDNVANVTESHREBEKLEPIPELIM  
 AOEETVQESITPOQVAKMEPCEGSKVYKLSPOSTTAVAVLVPSSKKQDQKMRFGQ  
 TVQVAVNSQAKVDVFAKSLPDCGPIENKXIVNGTGMFVLVGDPAITSEAKKALIA  
 QLPQEKELQPFVKSIDAIKNSIPTLANKN"  
 gene 9788..10783  
 /locus\_tag="VW20280"  
 9788..10783  
 /locus\_tag="VW20280"  
 /note="similar to ATP-grasp enzymes; COG1181"  
 /codon\_start=1  
 /transl\_table=11  
 /product="D-alanine-D-alanine ligase"  
 /protein\_id="AAO07245.1"  
 /db\_xref="GI:27358295"  
 /translation="MMRVVLLICGSGSESHESLISSTVLOOQGLIENVVLYKEIK  
 NGEWLDQERLYLIDHRTKSVKSDSEFNHSIHDFVPCITHPPEGTGTQSLFELAGI  
 PVLQGPSEKSSFNKTSKLYMDAIGLPPNPYVLTENNESHQLSREADRGKQKLE  
 VVAAGQSGVGCYSTKIEQLSDAIDKAFGFSHQVLYEKAVKPRELVEVSAIENAGQHL

Query Match	Best Local Similarity	64.9%;	Score 555.6;	DB 1;	Length 300883;
Matches	876;	Conservative	0;	Mismatches 459;	Indels 15; Gaps 3;
QY	25	TCGAGAAATTTGTTATTTGACGTTATTCGCGTTTCGTCGACCGCTTGGGGTTGACGATTGCG	84		
Db	118404	TCGATGAAAAAGAGATGAGACAGTTGGCGGTATTAACAAGTTGGTGGCTTGTCTTCAACG	1183		
QY	85	CCCGATGTCGGGGCAGTTGGCAGCTGAGAAAGAAAGAAATTTGAAATTCGGCTTCAATCAAG	144		
Db	118344	GTTGTACTCGGAGAGTCGGGAGCGTGAAMAAAGACATTAAAGTTGGCTTTATCAAA	1182		
QY	145	CTCACCGACATGCGCGCGCTGCGCGTGGCCGCCGAAAGAGCTCTTCAAGACGAGGCG	204		
Db	118284	TTGACCGACATGAGACCGTTGGCGCATGCTTATGAAAAAGATTTTGTGAGAGTGAAGCG	1182		
QY	205	CTGTTCTGTCACATCGAAGCGCAGCGCACTGAAAGGTGTGATGATGAAGGTCGTAAAT	264		
Db	118224	CTGTATGTGACGTGAAAGCGCAGGCCAACCTGGAAGGTGCTGCTGATCGGTGATTGAT	1181		
QY	265	GGCGAATCGACGGGCTCGCACATCTGGCGCCGGCGCCGTTAAGCGCCAGCGTTGGCTTC	324		
Db	118164	GGTGAACGTGATGCGCGGCACATCTGGCAGGCGCAACCACTGGGTCGCACCATCGGCAAT	1181		
QY	325	GGCACAAGCGCGATATCGAGGTGCGCTTACAGCATGCGCTTCAACGGCAGACGCAATTACG	384		
Db	118104	GGCACGCAAGCGCAAGTGATCAACCGCAATTCAATGATCTCAATGTGTAAGCATCAACC	1180		
QY	385	GTGTCGATGAAATCTGGCATTCAGATGAAGCGCAACTACCGCTGGA---AGCGGTAA	441		
Db	118044	GTGTCGAAACGATGTTTGGCAGCAATGAAAGCCCAATTTGGTCAAAAGCAGTGAACGCTAAA	1179		
QY	442	CCGGTGATCGGATCAAGGAGATTATCTAAAACCGGTGTCGAAAAATACAAAGCGGAA	501		



Db	117984	CCCGTTCAACCCCATCAAAAGCCGATGCCCTTAAACGTGTGTAAACAGTTTACCGCGACCA	1179925
Qy	502	GGCAACCGCTTTCAATATGCGATGACCTTCCGGCCGGATCTTCAACAATCAAACTGCGT	561
Db	117924	GGCAAAAGCGTTCAATATGCGATGAGTGTTCCTCCGTTTCAACGCATTAATCAAGAACTGC	117865
Qy	562	TACTGCTGCGGGCTGCGGTATTCATCTCGGTACTATTCGGCCGCGCA---GACATT	618
Db	117864	TACTGCTGCGGGCTGCGGTATTCATCTCGGTACTATTCGGCCGCGCA---GACATT	618
Qy	619	TCCGCGCAAAATGCGCGCAAGCCCTTGTGTGCTGATACCCCGCCGCGCAAAATGCCGTCC	678
Db	117804	AGTGCCTAAATTAATGCGGACGTCGTCTCAAGGTACGCGCCGCAACACAAATGCCAGCC	117745
Qy	679	ACGCTGGAAGCCGGCACCATTTTCGGTTATTTGCGTTCGCGAGCCGTGTGAACCAAGAGCC	738
Db	117744	ACCATGGAAGCCGGCACCATTCAAAGGCTATTTCGGTTGTGTGAACCGGTGAACCAAGAGCC	117685
Qy	729	GTGTTCAAGGGCATAGGCGTCGCGGTGATCACCGATGGAAGAATCTGTGAAGGACAGCCG	798
Db	117664	GTGTTTAAAGGGCATAGGCGTCGCGGTGATCACCGATGGAAGAATCTGTGAAGGACAGCCG	117625
Qy	799	GAAAAAGCTTCGCGCGTGACCAACAATGCGCGGAGAAATACCCCAACATTAATCGCG	858
Db	117624	GAGAAAGTATTTGGCGTTGSCACAAAGTTGGCGGGAATAATCCCAATACCATATCCG	117565
Qy	859	GTGACCAAGCGCTGATTTCCGCGCGCGATCTGCGTGTGAAGCGCGCAATTAACAAGAGCCG	918
Db	117564	GTGTCCTAAAGCTCTGATCCGTGCGCGGCGCATTTGTGTGATGAAGAACATTAACCGCAATCCG	117505
Qy	919	AAAGAAAGCATTCGAATGCTGGCGCAAAAACAATAGTGGTGTGATAGTGTGAAGTCTG	978
Db	117504	GCCGAAGCGGTGAACCTGCTTTCACAGCACCGAGTACGTGGCGCGCATCGGAAGTGATC	117445
Qy	979	GCGGCTACATGAATAGGCGCACTTTCGATTCGAATAAGAGATTAAGCGCGCATACCGGAC	1038
Db	117444	GCCATTTCCATATACCTGCGACGCTTTGATATCGAAAAAGCGATAGCGAACAAGTCCCGAT	117385
Qy	1039	TTCAACACTTCTTTCGCGCACGCGCCAGCTATTCGCTCTACAGCAGTGCAGTCTGTAT	1098
Db	117384	TTTATATGTTCTTTCGCTGCTACAAAGCGACCTATTCCTTACAGGAGCGCATTTGTAT	117325
Qy	1099	CTGACCCAGCTGAGCGCTGCGGGCATGATCAATGAATTCAAAACCGGACAATGTGTATCTG	1158
Db	117324	CTCAACAACAATGCGCGCTTGGGGAACAATCGAACAATCAGAAAAACCGGACTCAATGTATCATG	117265
Qy	1159	GATACCGGCAAGAAGTCTACAGCCCGGACATCTATCTGCGCGCGGCAAGGAATCGTGC	1218
Db	117264	GACATCGGCAAGCAGGTGATGACACAGATTTTATCAACGCGCGGCAAGGACATTAATC	117205
Qy	1219	GCAGAAAGCAAGGCCAAGCGCGAAGACTTCCCTGCG-----GATACCTGCATCAAG	1269
Db	117204	GAAAGAGGACATTTAGAGCGCCAGTATTTTCTGACTTTGCCAAGAAAGCCGTTTTCG	117145
Qy	1270	CCGTCGCAAAATTTTATATGACAAAGTGCCTTCGATGATCCAAACAAGCCCAACGATTAAC	1329
Db	117144	CCACGCAAAACCACTTCAATGATGATCCGCTATGACGCGCGGTAGCCCAATGTGTAT	117085
Qy	1330	CTGCGCAAGTTTTCGATAGTCTGAAGGC	1359
Db	117084	CTGCAAGCTTTCGATAGTCTGAAGGC	117055

RESULT 3	
LOCUS	AP005348/c
DEFINITION	AP005348 248650 bp DNA linear BCT 06-JUL-2004
ACCESSION	AP005348 BA000038
VERSION	AP005348.1 GI:37201116
KEYWORDS	
SOURCE	Vibrio vulnificus YJ016
ORGANISM	Vibrio vulnificus YJ016

REFERENCE	REFERENCE
AUTHORS	AUTHORS
TITLE	TITLE
JOURNAL	JOURNAL
PUBMED	PUBMED
REFERENCE	REFERENCE
TITLE	TITLE
JOURNAL	JOURNAL
1	Bacteria: Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio.
Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L., Liu, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P., Lee, C.T., Hor, L.I. and Tsai, S.F.	Chen, C.Y., Wu, K.M., Chang, Y.C., Chang, C.H., Tsai, H.C., Liao, T.L., Liu, Y.M., Chen, H.J., Shen, A.B., Li, J.C., Su, T.L., Shao, C.P., Lee, C.T., Hor, L.I. and Tsai, S.F.
Comparative genome analysis of <i>Vibrio vulnificus</i> , a marine pathogen	Comparative genome analysis of <i>Vibrio vulnificus</i> , a marine pathogen
Genome Res. 13 (12), 2577-2587 (2003)	Genome Res. 13 (12), 2577-2587 (2003)
14656965	14656965
2 (bases 1 to 248850)	2 (bases 1 to 248850)
Chen, C.Y., Wu, K.M. and Tsai, S.F.	Chen, C.Y., Wu, K.M. and Tsai, S.F.
Direct Submission	Direct Submission
Submitted (29-MAY-2002)	Submitted (29-MAY-2002)
Shih Feng Tsai, National Health Research	Shih Feng Tsai, National Health Research
Institutes, Division of Molecular and Genomic Medicine: 128,	Institutes, Division of Molecular and Genomic Medicine: 128,
Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan	Yen-Chiu-Yuan Road, Sec 2, Taipei 115, Taiwan
(E-mail: petals@nhi.org.tw, Tel:886-2-8146-1041,	(E-mail: petals@nhi.org.tw, Tel:886-2-8146-1041,
Fax:886-2-2789-0484)	Fax:886-2-2789-0484)
This sequence was determined by the Sequencing Core of the National	This sequence was determined by the Sequencing Core of the National
Yang-Ming University Genome Research Center (YMC;	Yang-Ming University Genome Research Center (YMC;
http://genome.ym.edu.tw).	http://genome.ym.edu.tw).
Location/Qualifiers	Location/Qualifiers
1. 248850	1. 248850
/organism="Vibrio vulnificus YJ016"	/organism="Vibrio vulnificus YJ016"
/mol_type="genomic DNA"	/mol_type="genomic DNA"
/strain="YJ016"	/strain="YJ016"
/db_xref="taxon:196600"	/db_xref="taxon:196600"
/chromosome="I1"	/chromosome="I1"
383. 1786	383. 1786
/gene="VVA0914"	/gene="VVA0914"
383. 1786	383. 1786
/gene="VVA0914"	/gene="VVA0914"
/codon_start=1	/codon_start=1
/transl_table=11	/transl_table=11
/product="anaerobic Ca-dicarboxylate transporter"	/product="anaerobic Ca-dicarboxylate transporter"
/protein_id="BAC96940.1"	/protein_id="BAC96940.1"
/db_xref="GI:37201117"	/db_xref="GI:37201117"
/translation="MKRVYFPCITRENETITSLRINNNLYEFLFLIMLYIGSR	/translation="MKRVYFPCITRENETITSLRINNNLYEFLFLIMLYIGSR
GGJGAGVSGIGLIVFVFKMPTSPPVVMIILAVVCAISLEAGGLKMLQVA	GGJGAGVSGIGLIVFVFKMPTSPPVVMIILAVVCAISLEAGGLKMLQVA
ERMLRKRKPVTLIAFVTVYTMFGTGLAVASIMPIGDVALKKGIRERPMMA	ERMLRKRKPVTLIAFVTVYTMFGTGLAVASIMPIGDVALKKGIRERPMMA
VASQIILATASPIAAVVYLAQSDINHETLLSILVTPALFGLMLSLYSMRG	VASQIILATASPIAAVVYLAQSDINHETLLSILVTPALFGLMLSLYSMRG
KELEDDPEVOARKDPVMEKRLINTATSLDEHLPAARNSVTLFSLITIVIMI	KELEDDPEVOARKDPVMEKRLINTATSLDEHLPAARNSVTLFSLITIVIMI
PERRTIGEGTKPIISMAVVIQMMMLCTGGIITLLATKTDPRSVPGVYFKSMAVAIAIF	PERRTIGEGTKPIISMAVVIQMMMLCTGGIITLLATKTDPRSVPGVYFKSMAVAIAIF
GIAMWSDITVYQVAMPQKSGIVEMVNNVPTFTALALIVSVVNSQAALVARMMLPVGL	GIAMWSDITVYQVAMPQKSGIVEMVNNVPTFTALALIVSVVNSQAALVARMMLPVGL
GLGLBEPALVGLMPALPVGAFYIPNYPSDIATVFNPDNTGTTKIGKWFNHSFMSVGLIG	GLGLBEPALVGLMPALPVGAFYIPNYPSDIATVFNPDNTGTTKIGKWFNHSFMSVGLIG
VIGACCCGYLGGMIIG"	VIGACCCGYLGGMIIG"
1858. 1891	1858. 1891
/gene="rbc2T08"	/gene="rbc2T08"
1858. 1891	1858. 1891
/note="2 units of 17 bp in-tandem repeat"	/note="2 units of 17 bp in-tandem repeat"
/rpt_type=tandem	/rpt_type=tandem
/rpt_unit="ttcattgcgcctaac"	/rpt_unit="ttcattgcgcctaac"
complement (1905. 3185)	complement (1905. 3185)
/gene="VVA0915"	/gene="VVA0915"
complement (1905. 3185)	complement (1905. 3185)
/gene="VVA0915"	/gene="VVA0915"
/note="identified by Genemark and Glimmer2"	/note="identified by Genemark and Glimmer2"
/codon_start=1	/codon_start=1
/transl_table=11	/transl_table=11
/product="conserved hypothetical protein"	/product="conserved hypothetical protein"
/protein_id="BAC96941.1"	/protein_id="BAC96941.1"
/db_xref="GI:37201118"	/db_xref="GI:37201118"
/translation="MMKRRILTLTLTQWLMOPAVILAAGTNNVSOYTVARVWMAOQALAD	/translation="MMKRRILTLTLTQWLMOPAVILAAGTNNVSOYTVARVWMAOQALAD
NNLALAIKLNSAEVSRPDDVALSLKLGIVYWAQOFRTASISLEKALSFLTFBE	NNLALAIKLNSAEVSRPDDVALSLKLGIVYWAQOFRTASISLEKALSFLTFBE
EBWVTVRMLTOLITLQGDYKKAIPOLNKLKQVPEBSAABIRVYLAQAQSVENNR	EBWVTVRMLTOLITLQGDYKKAIPOLNKLKQVPEBSAABIRVYLAQAQSVENNR
TVATAGFQPRVREKPPPTTISLALAHVQLEOKSVYIPAKQIILAEPSKNNWLAAY	TVATAGFQPRVREKPPPTTISLALAHVQLEOKSVYIPAKQIILAEPSKNNWLAAY
NALINIGQSKALDMTLIELQGLBNKEKETSIALVIAAQIGYQFAQOTLARTESQ	NALINIGQSKALDMTLIELQGLBNKEKETSIALVIAAQIGYQFAQOTLARTESQ
QDQKILKIQAVYQAKKEWQALIEQRLAKQESQYQWEVALLLEQLOVQVLLISA	QDQKILKIQAVYQAKKEWQALIEQRLAKQESQYQWEVALLLEQLOVQVLLISA
EBMNPGRQFPAQLKINALYRLERIDQAAVQAQKDALSTKOTQWMLKFLSHKLS	EBMNPGRQFPAQLKINALYRLERIDQAAVQAQKDALSTKOTQWMLKFLSHKLS
ESLDLAKSGQVSEERHLLTOQRNVADKSDSS"	ESLDLAKSGQVSEERHLLTOQRNVADKSDSS"
complement (3182. 3820)	complement (3182. 3820)
/gene="VVA0916"	/gene="VVA0916"
complement (3182. 3820)	complement (3182. 3820)
/gene="VVA0916"	/gene="VVA0916"

```

/codon_start=1
/transl_table=11
/product="periplasmic protein TonB2"
/protein_id="BAC96942.1"
/db_xref="GI:37201119"
/translation="MRRLAFPPVALALVFALFSAMALLVNSGKQALDQTSLSFNM
VAVEQDEVTQRQSRVPPPEIPEPQEPETPKQTPQMVQTSOVAETPIDRFPVNSG
LAVSIPATVSAPTONASNTTKQMLPLSRVAVYPAKAKKRIEGVYLKFDIDENGR
KNIEVEAQPAPFEENMDAIKRMRYQPIYDGAQITTYGTTKIEFMAQ"
/gene="VVA0917"
/complement(3820..4224)
/gene="VVA0917"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbD2"
/protein_id="BAC96943.1"
/db_xref="GI:37201120"
/translation="MRIGRQSGDEAQIDLTSMDIVFIMLIFPIVTSFYRESGVE
VNRPOANYSACKDAGIFVAITANDIYIDKRMVDVERVQASIERLLLEQPPASLVIO
ADEYANGVYVYMDAKAAGKNTALAERS"
/gene="VVA0918"
/complement(4221..4787)
/gene="VVA0918"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbD2"
/protein_id="BAC96944.1"
/db_xref="GI:37201121"
/translation="MIMDLHSDIWPBELLTSSSMWNALYFMOQGEVLMMLAAVAVI
PMLTIEKLFIYCAVPRRAETLQKQMRHDLISWFAISQREAMLDHQQLPFHLA
FKVVAICPMIGLIGTVGMISVDFVMAVTOGSSNPKMAAGISLATLTPMAGMVAAL
VGMFYHARIVYKCONREVLTEOSLRDS"
/gene="VVA0919"
/complement(4784..6151)
/gene="VVA0919"
/codon_start=1
/transl_table=11
/product="biopolymer transport protein ExbD-related
protein"
/protein_id="BAC96945.1"
/db_xref="GI:37201122"
/translation="MNKCRLLISLALVSLSPSPAGTOLVPRKQTESROQAOENORE
NAFLSQEQLAKLEKQKSTLEBKNALESQSDNVTLSHBEELRYTSGISGE
LPGVROHAKELKADINDAVTAIGQTEQVETVNOIVAAQSLPIQAKTLHVMQOQI
VASGEMRALQVNLINAEQAVEPTDVNALFALFALQGFVNNKSGGFIIPYAKLPG
APTLDLRSQOHAISMVPLVIDPRTGVLEQVYHSPSLADRFAGGVGKVFIPALLAVG
LVIALYRGILLTTSKRILOLKOPNQPNPLGRLAVYDQKPRSVALELRIET
IYDEORTGKGLSMLKLAALAPMLGLIGTVGMIEETPQVITQFGNGDPKWAQISGM
ALVTYVGLIVAAEMPLAHNVLSGMAEIRSLIEKQIGLVAQAEQULPSLPTMDK
VGNAA"
/gene="VVA0920"
/complement(6152..6919)
/gene="VVA0920"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC96946.1"
/db_xref="GI:37201123"
/translation="MSGYKAGAAALLIMAVSFGAMSNDEQAQKIQNRRTTASQSV
INASASQLOLAHEITLKEVNRBELYRHLTSMANQERKSSDAEKQIDQIQTRO
GIVPLMYDMLDTLKTITINDMPKMQOQREIRLEKQMWQADVADAEFRILHMAVS
IELDYSSKLSAYOAPITTLADGQIEVEQVHLGVSILATLINANOFTWYDQQRHLS
LTTIPSTDVNRAFSDIAACKVAPAMLMMPALATAEAK"
/gene="VVA0921"
/complement(7182..7670)
/gene="VVA0921"
/codon_start=1

```

```

/transl_table=11
/product="uncharacterized conserved protein"
/protein_id="BAC96947.1"
/db_xref="GI:37201124"
/translation="MRWTSRYALFVVDVONITYTCKERYRHFYNNRQSATQDKQV
VVAHAYALSSDPCGRQFPHILIRGIFGVKPKPFIORDGSAKGMVDGLADAEHA
PMVEVILLSDGDFEILVERIQKFGKVTYVGVPLTAQNLQAVVDKYPIDELL
LP"
/gene="VVA0922"
/complement(7796..8227)
/gene="VVA0922"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96948.1"
/db_xref="GI:37201125"
/translation="MRWYSIVLILFSLAHSVPADSEKQHRSIMFPAPSHQVYQOF
LLETIMESCSLOERDIYTLVTKEDSGSTVPVNAQDFDLSVNAFVDIPKQHTILVG
KDGREKLRMGKRTMTLLKVIDQMPLRQEMQNRPSRCAT"
/gene="VVA0924"
/complement(8234..9076)
/gene="VVA0924"
/codon_start=1
/transl_table=11
/product="uroporphyrinogen-III methylase"
/protein_id="BAC96949.1"
/db_xref="GI:37201126"
/translation="MLIKSKHFPREALSGTLHQIYISIGSTMILTSLSSLOQFVS
LVGAGQPDILTIVGYVIVIOAEVVDVDRISPEITLALANEKAMIYVGKLDPRCV
POEINNOILVRAKLEGRVRLKGGSDTIFRSGEELBELHGVRFVDPITTAAG
ATAVAGIPLTRHDQASQVPTTGHIOKRGKLEMPSLAONNTIYFWGLQCRVIAE
KLENGLDASTSCALIENTRBEQVRLTGSILTQPELTAQISPLIYVGSVTSLHDK
LKW"
/gene="VVA0923"
/complement(8255..8395)
/gene="VVA0923"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC96950.1"
/db_xref="GI:37201127"

```

Query Match 38.9%; Score 541.2; DB 1; Length 248850;  
Best Local Similarity 64.2%; Pred. No. 5,7e-94;  
Matches 867; Conservative 0; Mismatches 468; Indels 15; Gaps 3;

QY	25	TCGAAGAAATGTTATTCGCTATCGGCTTCGCTAGCCGTTGGGGTTTGACGATGGC	84
DB	46865	TCGATGAAAAAGATGACAGTTGGCGTATTAAACAGTTGGCTTGCCTCCACCG	46806
QY	85	CCCGATGTGGGGCGATTGGCAAGCTGGAAAAAGAAAGATTGAATTCGGCTTCAACG	144
DB	46805	GTTGTACTCGCAGAGTGGCGAGGCTGAAAAAGAGACTTAAAGTTTGGCTTATCAAA	46746
QY	145	CTCACCAGCATGGCGCGCTGGCGGTGGCCGCCGAAAAAGCTTCTTGAGAGCAGAGGC	204
DB	46745	TTGACCGACATGGCACCGTTGGCGATCGCTATTGAAAAAGATTTTGGAGATGAAGGG	46686
QY	205	CTGTCCTGTCGCACTGGAAGCCGAGCCCACTGGAAGTGTGATGATAGGTCCTGAT	264
DB	46685	CTGTATGTGACGCTTGAAGCCGAGCCCACTGGAAGTGTGCTGATCGCGTGAATGAT	46626
QY	265	GGCGAATGAGACGGCTCCCACTGTGGCGCGCGCGCTTAGCGGCCAGCGTTGGCTTC	324
DB	46625	GGTGAATCGATGTGTGCCCACTGTGGCAAGCCCAACCACTGGGTGCCACCACTGGCAAT	46566
QY	325	GGCACCAAGCCGATATGAGAGTGCCTTCAAGCATGGGCTTCAACGCGCAAGCATTACG	384







access to supplementary information.

FEATURES	source
Location/Qualifiers	1..311050
/organism="Rhodospirillum rubrum SH 1"	
/mol_type="genomic DNA"	
/strain="1"	
/db_xref="taxon:243090"	
/complement(3..182)	
/locus_tag="RB2"	
/complement(3..182)	
/locus_tag="RB2"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein-signal peptide and transmembrane prediction"	
/protein_id="CAD71330.1"	
/db_xref="GI:32442812"	
/translation="MSRLFFSALFAFVSVFTVGGCGGNEVIEDTRSPAEVQEOED YEKOMEIWDSATNS"	
39..260	
/locus_tag="RB3"	
39..260	
/locus_tag="RB3"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein-signal peptide and transmembrane prediction"	
/protein_id="CAD71331.1"	
/db_xref="GI:32442813"	
/translation="MLFVIFLLHFGWRAVGFDFVAASATNSEDRNAERKQGEK OSRHYCSMDARNECEKLVRELLILRLISP"	
/complement(399..1514)	
/locus_tag="RB5"	
/complement(399..1514)	
/locus_tag="RB5"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein-signal peptide and transmembrane prediction"	
/protein_id="CAD71332.1"	
/db_xref="GI:32442814"	
/translation="MTKRLPTRQRLVLAAGLITPLAASTAKAGFQWETGSPY YEDDAVYDSEWFDGNDYPTDEAIGRMDEYDDEALTSDDNDIDMTWKDGXY ADTSSDTFVYDHYDGYDGSNNNDGNYEYTSYVYDGVYDAFASVYDTGGD IYDDATYISFNSAEDNSOKTAKKADONASAKSOFSTGTTAKKVTATNTNV VELQNSBEKSMIADLGPADQIDRMPQNDTITVYSGSPKADTIVLSRSIEHNGQ MVERSGRTSGTISTKTVDVGGQRLAKITDGGKMLVLDGRKODLDTNTEGS QVTNGPAVQVDRVWLATKVTLDGDTIEISRMOK"	
1548..1706	
/locus_tag="RB7"	
1548..1706	
/locus_tag="RB7"	
/codon_start=1	
/transl_table=11	
/product="hypothetical protein"	
/protein_id="CAD71333.1"	
/db_xref="GI:32442815"	
/translation="MRSRSQSVVOETNGKVDKADIGAAVVAQPNRREGKLSVAG RGGKQNVQ"	
/complement(1733..5263)	
/locus_tag="RB8"	
/complement(1733..5263)	
/locus_tag="RB8"	
/function="DNA synthesis and replication; transcription"	
/note="Best DB hits: BLAST: p1r:D83545; probable helicase P07099 [imported] - Pseudomonas; E=7e-95 emb1:CAA67095.1; (X98455) SNF [Bacillus cereus]; E=8e-94 p1r:C72027; sw1nf family helicase 2 - Chlamydomonas reinhardtii; E=4e-87 COG: P00799; COG0553 Superfamily II DNA/RNA helicases; SNF2 family; E=7e-96 PPM: P02803; Thiolase, C-terminal domain; E=0.11 Pf00176; SNF2 and other N-terminal domain; E=3.9e-64 Pf00271; Helicase conserved C-terminal domain;	

QY	79	ATTGGCCCGCATGTCGGGCGAGTGGCAAGCTGGAAGAAAGAAATTGGAATTCGGCTTC	138
DB	184764	AATCCAACTGATGCTGCTGCATATGTTGACATTTGGAGAAATGCGATTGACATTCGCTTC	184822
QY	139	ATCAAGCTCACCAGCATGACCGCCGCGTGGCGCGCCGCAAAAAGCGCTTCTTCAGAGAC	198
DB	184824	ATCAAACTGACGACGATGCGCTCTTGGTGATTCGCAAAAGAAAGGCTACTTCAGAGAC	184883
QY	199	GAGGGCTGTTCTGTCGAACCTGGAAGCGCAAGCGCAACTGGAAGCTGTGATGATGAGGCTC	258

E=1.2e-24

/codon\_start=1

/transl\_table=11

/product="helicase"

/protein\_id="CAD71334.1"

/db\_xref="GI:32442816"

/translation="MDWIAEDQSLGILMGSAVERLVSDHOGFDSRAAMLTVKTPT  
LTQDNGVAFDFQVYVNGREKQTVALELTPQVEDDPEGVELESCCGCYAFLRDN  
SCPHIAAAMWLOEQIARRNGNGOLIEPLGKADVAAGKQFVELIRRAKNSRAP  
AEDAEPRLQWRVNRVNRASOTAPLITTPPEQOKRGKRGREVRSEDLIDDLIT  
DPLDSKIASLTSVPTHGVNDNYALFELRLMAGBENYAMDONALPVEVFCQSLLT  
LQPEVEEQSKRPKASQTDGASKKTAATRRSSSKRSKGSADLELGNATATMT  
PESLDSIEAPSAOGLIPDDIIPDEKVEVOFOPRLSYDGLIDIVQCEIIVGASPV  
EPAIILADADBDRIVLCIADPAALIGILOFLRGDYRPLTDHATASRLALDASIDS  
LIRVELPOLAGPIVAPCELVQWRPARGAKLKSILVYESRLDSILIPGAPELVP  
CLARGTRIQORDKAELESAAKTYORGSRLSHESTMEVVALTDHAAIDLGLHD  
AGDSERIVPPEGETFVRVREGITPKSLRIKIDSDPDMGNTGTVTVQGVDFLOQLM  
AVREDRQLVRVGDERRPAKISAFRRRLTQIDTVLSEKQURLAGAAVPLADVLGDD  
VTLLETTAAMHQAIERLDSIQDMSDPRPEGLDATALYQLDGRMTARLSRMVGGVLA  
DDMGLKTVQTLGILVERGPVGPALVAVPVGSEWVRAEAKFTPALTGKLRDCDRD  
EVIKTAGNDLIVSYOLQORDAKRFAPRGWHTLVLDRAOFIKNSQTSOAIPTIOA  
DMRGLSTPIENHGLGELMSLFRILSPGLLSMQRFRFRAPPIFRHDDARRSELAR  
LVRPFLVRKTDKYLKEPLRTEITLRKELSAPEKLYEDAFALALALALMGGAPSH  
EGRRIQTLISWLRRLQLACHPSLVPEPSWKCTISSQLFLSLVEELREGDRALVFSQ  
FVKHLSVRAALDERGISYQYLDGATPSHEKORSRRVADQNGSGDLFLSLKAGTGILN  
LTAADYVHLDPMMNPVAVEDQATBRAHIGERAVTVYRLVARTIEBQIILQHADKR  
ELVAGILGTBAAKLOTODLIDLKAEWG"

complement(5458..5850)

/locus\_tag="RB13"

complement(5458..5850)

/locus\_tag="RB13"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAD71335.1"

/db\_xref="GI:32442817"

/translation="MHLPHGNQRFKRTKYNRETSETLLRTLATETVRLTPVAHPF  
PRSFEIEMETRESSDPQLSKDCPKIPSTIIIGNKVCSPSVIPKEIPLHTTCQEPPLSNR  
AGNRRVDIKRGRVGFGRVPLRLRLSP"

5836..7386

/locus\_tag="RB14"

5836..7386

/locus\_tag="RB14"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="CAD71336.1"

/db\_xref="GI:32442818"

/translation="MGQMFDPVPPAANEFESTREDAAYLCIGVPMQNSNSNGR  
FTLRGVDTSAKLEFLTCPIPSIGYRQSLTCSLRCDRSHRLFVEIARSGCYVRVQAD  
VMORGALVLSRFDPLFRGTRFLDAACQAPRDPSCVDALAEALAEIASHNSNGLD  
FSTOSVAYARRSRERLSTMLAVGLPRLPAIEVYADMGSGASNSIDEOATTTAFNA  
AAVIRISNGEIEITDSGRPVYEPTEDEALTCACQMGARVYGVYIDRKVFLPMULTID  
DFDKLMSTWTSFVEKTVQFRKRVNLMAAGLAVAGPLGFDDEGVNMFSGILDTIR  
RCDPVTPLVMGLIDQPCGEYIARDEQGSPLHFAALARSGLAGIQLNFRGFDGD  
THPSAVFGCMIDRMGTNLNAPLWVQLSVAGAPGDAARAPITQRPISPEINGSDPA  
WAKQIEFVQPLIEITLSKQIVHNVMDGMSDQPHNTPRHAGVIDAKGRPRMPLAYLT  
KIREMLI"



Db 184884 GAAGACTGACGTCACTTTGAAACCACTGACAGTGAAGATCCTTCTGACAAAGTC 184943  
 QY 259 GTGAAATGGGAAATGGACGGCTGACATGCTGGCCGGCCGCTTAAAGCGCCAGCGTT 318  
 Db 184944 ATCAATGGGCAATTCACGGCGGACACATGTTGGCCGCGCAACCGATCGTGCACGATC 185003  
 QY 319 GGCTTCGGGACCAAGGCGGATATCGAGGTGCGGTTCAGATGGGCTTCAACGGCAACGG 378  
 Db 185004 GGTGTGGAACCAAGTCGCCAATCTGACCGCTTACGCTTGACATCAACGGCAATGGC 185063  
 QY 379 ATTACGGTGTCCAATGAATCTGGCATCAGATGAGCCGCAACATACC---GCTGAAGGC 435  
 Db 185064 ATCAAGGTGACGAACGAAGTTTGGGCAACAGATCAGAGAAAGATCCTGAAGTGAAGAG 185123  
 QY 436 GGTAAACGGGTGATTCGATCAAGGCAATTTCTAAACCGGTCTCGAAAATAC--- 492  
 Db 185124 CCCACACCAAAATCTCCATCAGCGCGCGCAGCTGAAAACCGATTCAGAGATTACCTG 185183  
 QY 493 AAAGCCGAAGGCAAGCGCTTCAATATGGGATGACCTTCCCGCGCGGATCTCAACATC 552  
 Db 185184 CAAAGACGGGGGCAACCATTTCCGATGGGATGATGTTCTGTTAGCACTCAAAATTAC 185243  
 QY 553 AAATCGCTTATCTGCTGGCGGCTGGCGGTATCAATCTGGCTACTATTGCGCGCGCAG 612  
 Db 185244 GAGATTCGATATTGGCTGGCGCGGTCAAGGATTCATCCGCGCATATCACCGAATC---T 185300  
 QY 613 GACATTTCCGGCAAAATCGCGCGAAGCCTTGTCTGCTGATCCCGCGCGCAATG 672  
 Db 185301 GACATCAAAAGGCTTCAACGACGACAGTCAATTTGTCCACGATGTCACGCTCAAAAG 185360  
 QY 673 CCGTCAGCTGGAAAGCGGACCATTTTGGTTATTTGGTGGGCGAGCGGAGACAG 732  
 Db 185361 CTTCAAGACTGGAAAGCGGACATCTCAAAAGTTACTTGGGCGCAACCGTGAACCA 185420  
 QY 733 CAGCGCTGTCAAGGCGATAGCGCTGCGGTGATCAGCATGAGAACTTGAAGAGAC 792  
 Db 185421 AAAGCGGTGTGATCTGGCATCGGTGTTCCAGTCAACGACATACAGACATCTGAAGAAC 185480  
 QY 793 ACCCGGAAAAAGTCTTGGCGGTGACCAAAACAAATGGGCGGAGAAATACCCCAACCTAT 852  
 Db 185481 AACCTGAAAAAGTCTTGGCGGTGACCGGAGATGGGCGGAGAAATACCAAACTCAC 185540  
 QY 853 CTGGCGGTGACCAAGGCGGTGATTTGGGCGCGGATGCTGGAGCGCGCAATTAACAG 912  
 Db 185541 TTGGCGGTGATCAAGGCTGTGATCCGTCTGGCAAAATGGCTGATGCCCAATTAACAG 185600  
 QY 913 -----AACCGAAGGAGCCATCGAAATCTGGCGCAAAACAAATACGTCGT 960  
 Db 185601 GGCAAATTGTCATATGCGAAGAGGCGGTGAAATCTCAGCGCGCAAAATTAATGTCGT 185660  
 QY 961 GCTGACGTGAAGTGTGCGCGGTGACATGAACGGCACTTTTGAATACAAAAAGAGAT 1020  
 Db 185661 GCGGAGAAAGAGTGTGATGCAACTGATGATGGGACGTTTGTCTTCAATTCACCGAC 185720  
 QY 1021 AAACGGCGCTTACCGGACTTCAACACTTCTTTCGCAAGCGCGCGCACTATCCGTCTAC 1080  
 Db 185721 GTTTCGGAATGCCGACTTCAACGTTCTTCAACAGAAAGCAAGTCAACCGGACATAC 185780  
 QY 1081 AGCAGTGAAGTGTGATCTGACCCAGCTGAGCGCTGGGCGCATGATCAATGAATTCAA 1140  
 Db 185781 AGCGATGATGATTTGGTTCTTCAACAGATGCGACGCTGGGGAACAAATCAAGATTCAA 185840  
 QY 1141 CCGAACAATGATCTGATCTGATACCGGCAAGAAAGTCTTACCGCGCGGACATCTATCTGCGC 1200  
 Db 185841 CCCGCAATGATGATACCGCAAAACGCGCAAGAAAGATCTTACAGCTGATGATCTATGCGCAA 185900  
 QY 1201 GCGGCAAGAACTGATGCGCAAGGCAAGGCGCAAGCGGCAAGACTTCCCTCTG 1253  
 Db 185901 GCGCGCAAGCTTCTAATACGCAAGGCAAACTGATCCAAATGAATCCGCGC 185953

RESULT 6  
 AE017282\_06/c

WPCOMMENT  
 Sequence split into 33 fragments. Locus AE017282 Accession AE017282  
 Fragment Name Begin End  
 AE017282\_00 1 110000  
 AE017282\_01 100001 210000  
 AE017282\_02 200001 310000  
 AE017282\_03 300001 410000  
 AE017282\_04 400001 510000  
 AE017282\_05 500001 610000  
 AE017282\_06 600001 710000  
 AE017282\_07 700001 810000  
 AE017282\_08 800001 910000  
 AE017282\_09 900001 1010000  
 AE017282\_10 1000001 1110000  
 AE017282\_11 1100001 1210000  
 AE017282\_12 1200001 1310000  
 AE017282\_13 1300001 1410000  
 AE017282\_14 1400001 1510000  
 AE017282\_15 1500001 1610000  
 AE017282\_16 1600001 1710000  
 AE017282\_17 1700001 1810000  
 AE017282\_18 1800001 1910000  
 AE017282\_19 1900001 2010000  
 AE017282\_20 2000001 2110000  
 AE017282\_21 2100001 2210000  
 AE017282\_22 2200001 2310000  
 AE017282\_23 2300001 2410000  
 AE017282\_24 2400001 2510000  
 AE017282\_25 2500001 2610000  
 AE017282\_26 2600001 2710000  
 AE017282\_27 2700001 2810000  
 AE017282\_28 2800001 2910000  
 AE017282\_29 2900001 3010000  
 AE017282\_30 3000001 3110000  
 AE017282\_31 3100001 3210000  
 AE017282\_32 3200001 3304554  
 Continuation (7 of 33) of AE017282 from base 600001 (AE017282 Methylococcus capsulatus  
 Query Match 12.3%; Score 171.2; DB 1; Length 110000;  
 Best Local Similarity 53.3%; Pred. No. 8.4e-23;  
 Matches 542; Conservative 0; Mismatches 408; Indels 66; Gaps 6;  
 QY 125 TGAATTCGGCTTCTCAAGCTCACCGACATGCGCCGCTGGCGGTGCGCGGAAAAAG 184  
 Db 18318 TGAACATGGGTTTCAATCTCTTCAAGCTGCGATCGGCTCCGTGATCTGCGGAAAAAG 18259  
 QY 185 GCTTCTGAGGACAGGAGCGCTGTTCTGTCACATGGAAGCGAGGCAACTGGAAGGTGG 244  
 Db 18258 GTTACTTCGGGAGACAGCGGCTGAGACGTACACTGTTCGCGCAAGCCGCTTGGGCGCCA 18199  
 QY 245 TGATGATAGGGTGTGATGATGCGAATGCGAGCTCGCACTGCTGGCGCGCGCGCT 304  
 Db 18198 TCCGGAAGAGTTCGAGACCGGCTGCTGGAACGGGCCCAATGCTCGGCGCGGATGCCGA 18139  
 QY 305 TAGCGGCAAGCTTGGCTT---CGGACCAAGGCGGATATCGAGTCCGCTTACAGATG 361  
 Db 18138 TGGCCACCACTCTCGGCTGGGGGGGGGATACGGAAGGCTACATTCGCGCGATGTCGTGG 18079  
 QY 362 GCTTCAAGGCAAGCGGATTAAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 421  
 Db 18078 ATCTGAATGGAAGACCGCTGACGATGTCTCAACGCTCTGATC-----GCGGCTCA 18028  
 QY 422 TACCGGTGAAGGCGGTAAACCGGTGATCCGATCAAGGCAAGATTAATCAAAACCGGTG 481  
 Db 18027 CCGCGGTGGCGCGGTGAGCTGCGACGCGCGCGGTTTCGCGCGGCTCTCGCGGCTT 17968  
 QY 482 TCGAAAAATACAAACCGGAGCAAGCGGCTTCAATATGCGATGACCTTCCGCGCGGAT 541  
 Db 17967 TGTGGAACAGGAGCGGAGGCTTGTGCGGTTCGCGCGATATTCCTTCTTC-----T 17914  
 QY 542 CTCAACAATCAAACTGCGTTACTGGCTGCGCGGTGCGGATCAATCTGGCTACTATTT 601  
 Db 17913 CGCACAATCACTGCTGCTGCTATTTGGCTGGCTGGCGCGGATGATC----- 17865

```

QY      602  CGCCGCCGAGGACATTTCGGCCAAATGGCGGACAGACGCTTGTGTGTGACGCCGCG 661
Db      17864  -----CGACCGCCACGTCGGCAGACGCGTCATTTCGCG 17833
QY      662  CGCCGCAATGCGGTCACGCTGGAAGCCGCGACCAATTTTCGTTATTCGTCGCGGAGC 721
Db      17832  CGCCGAGATGCGACGTTGCTGCGCGCGGGGAGATTTCCGCGTACTCGCGGGGTGAAC 17773
QY      722  CGTGAACACGACGCGGCTGTTCAAGGCGATAGCGCTGCCGCTGATCAACGATGAAGAAC 781
Db      17772  CCGTGAATCCCGTGTGCGGCGGCGGATCGCGCTGACTGATCAACCGGTCAATGAAG 17713
QY      782  TCTGGAAGGACGCGCGGAAAAAGTCTTGGGGGTGACCAATTCGGGGGAGAAATACC 841
Db      17712  TCTGGAACACAGTCCCGAAAAAGTCTGGGCGTCAACAGGACTGGGGCGAAACCAATC 17653
QY      842  CCAACACCTATCTGTGCGGTGACCAAGCGCTGATTCGGGCGCGCATCTGTGACGCGC 901
Db      17652  CCAATATCCTTGGGAGCGCTCTCATGCGCCCTGTGGAACCGCGCAATGGCT-----CG 17539
QY      902  ACAATTAACAAGACCGCAAGAAAGCCATGAAATGCTGCGCAAAAAACAATACGTGCGTG 961
Db      17539  ATCGGCGCCGAAAAACCGCGGAAAGCGCGCCCTCTCTCGCAATCCGCACTATGTGCGCG 17539
QY      962  CTGACGTGGAAGTGTGCGGCGGTAGCATGAACGCACTTTCGATTCGAAAAAGACGATA 1021
Db      17538  TGGCGCGGAGATCTTCCAGATGCGATGACGGGTTCTTTCAATATGCGCGGCGGAGAC 17479
QY      1022  AACGCG---CGCTACCGGACTTCAACACCTTTTTCGCCACGCGCGCAGCTATCGGTCT 1078
Db      17478  CGGCGCGTCCCTCCCGGACTTCAACGATGCTTCAACGCGCTATGCGCGACCTTCCCTGCG 17419
QY      1079  ACAGCAGTCACTGTGTATCTGACCCAGCTGAGGCGTGGGGCATGATCAATGAA 1134
Db      17418  GTTCCACGCGCTGTGTCTTGAACCAAGATGCTGCGTGGGGGAGATCGCCGAA 17363

RESULT 7
AE008293      10676 bp  DNA  linear  BCT 28-MAY-2004
LOCUS      Agrobacterium tumefaciens str. C58
DEFINITION  187 of the complete sequence.
ACCESSION  AE008293 AE007870
VERSION    AE008293.1 GI:15159385
KEYWORDS
SOURCE
ORGANISM
REFERENCE
  1 (bases 1 to 10676)
  Hinkle,G., Slater,S.C. and Goodner,B.
  Complete Genome Sequence of Agrobacterium tumefaciens C58
  (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
  Disease in Plants
  Unpublished
JOURNAL
  2 (bases 1 to 10676)
  Hinkle,G., Slater,S.C. and Goodner,B.
  Direct Submission
  Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney
  Street, Cambridge, MA 02139, USA
COMMENT
  Approximately 800 bp of telomeric sequence missing from the left
  end of the chromosome and 200 bp missing from the right end.
FEATURES
  source
    1..10676
    /organism="Agrobacterium tumefaciens str. C58"
    /mol_type="genomic DNA"
    /strain="C58"
    /isolate="Cereon"
    /db_xref="taxon:176299"
    /complement(93..1352)
    /gene="AGR_L_1874"
    /complement(93..1352)
  gene
    CDS
      complement(93..1352)

```

```

/gene="AGR_L_1874"
/note="HYPOTHEICAL 40.1 KDA PROTEIN IN SIPU-KIP1
INTERGENIC REGION"
/codon_start=1
/transl_table=1
/product="AGR_L_1874p"
/protein_id="AAK89510.1"
/db_xref="GI:15159386"
/translation="MERKSMEOKKLTPTADLXSKRASLLGATFMAATSAIGPFI
QTATFTDGAFAFGILASTLIDFVQDNIWRTVTLTKRASDITANAIKPGAFIA
ILVITGGFFNIGNIGTGLNLHFGDLPKIGGAIISAFISIAIFVSRAGLAVRFI
IFAGILMTVLTLVYA FVSAPVGDAPROTLPDPTINFATITTVIGGTGGYITVGA
RLDRGMVGIENLPVNRALTLGIAVTGIMRVTLFLLGVASGVVDITSGOANPA
AQRPSAGDIFRLFGIIFMAAITSVIGAAYSVSPFPVKDMSRRAMAVF
IANSVCLTLTTPPAMLVFVGLNGVILFGLSIFLPAWKEDLMGHRYSILL
GLGVLTCLTVMYMGKSGTIFGLGL"
1517..2266
/gene="AGR_L_1876"
1517..2266
/gene="AGR_L_1876"
/note="(AUJ3213) putative GntR-family transcriptional
regulator."
/codon_start=1
/transl_table=1
/product="AGR_L_1876p"
/protein_id="AAK89511.1"
/db_xref="GI:15159387"
/translation="MCTGQMPCEMPPTPEPMHCHSERSAMTETIGLPLADRLAK
GIRTLISGELAPQORLEAFAFSRFDPSRLSLDEAFLLTKDGLRREARGVAV
PTMSVIDIYFVRMVEBCLOAMHKHPAIAKMAVAEDAKIRREEDMDVGVSAM
KFHAAYVDLQDSARISDFYERIAELRLCFGLKDPEQLHSPFVDMNSKILMEBG
AOBANEMELVNLISERTVLCALRAVHTS"
2341..2955
/gene="AGR_L_1879"
2341..2955
/gene="AGR_L_1879"
/note="response regulator nastr - Azobacter vinelandii"
/codon_start=1
/transl_table=1
/product="AGR_L_1879p"
/protein_id="AAK89512.1"
/db_xref="GI:15159388"
/translation="MLRCVINADMTLRDLTTLVIDENARASIIIEGREGAYQRTV
IHEVNGARTIETLOPYIFIDLPNPRDMWHLFOLRTVGRPIAMVVDSDTASIS
AAVAGVAAVVDLTKKRVKPTIDMAVSRNARSRLORLADKALSBEKKVIERAK
GILMKRGLSEEPFALIRQTPANKEKKISITIAOSVTTAAGLIM"
2991..4286
/gene="AGR_L_1881g1"
2991..4286
/gene="AGR_L_1881g1"
/note="hypothetical protein"
/codon_start=1
/transl_table=1
/product="AGR_L_1881g1p"
/protein_id="AAK89513.1"
/db_xref="GI:15159389"
/translation="MAAPEKTAGSKGATOWRDAPALVGSROKILIRAGTPIVDASV
LIAABREFADREGILIDVVDVSMVNRDLAROPIDARMSPMVPSNLGASNP
SPITTPSLRGSNALITSTRLEAFMKALTLSTAGLLENARALKVLDMRAGEA
PPTIGMTYPPFSHNYEFRYMLAAGIHFDHDKVAVPPMTSDALAGAIIDFCVGA
PWNIVAAERGVIRVAAKODLWPAPEKVIQMRBEMASQOETVGRLLTALDAASWC
DLANDHDLSCALADPRYIGAPQSIIRVLVAGERSIDSGNRVIEKFTYGDHIANV
PROGOSIMYSGMIRMGQAEISFGVNAALSAVRPDYRALGSGKAGADDIRIEGQ
DEGRFVDGFPFVDADTAGVNSFVARTSVVSAFSGEA"
4544..5959
/gene="AGR_L_1886"
4544..5959
/gene="AGR_L_1886"
/note="nitrate-binding protein ntrA precursor, periplasmic
- osillatoriacean cyanobacterium"
/codon_start=1
/transl_table=1
/product="AGR_L_1886p"

```

```

/protein_id="AAK89514.1"
/db_xref="GI:15159390"
/translation="MLPGTTPVPKAQTPEGVFIEFFPKRRRHPRKRWINKGMKMKK
ITSGIVSRRTILKTATTAALVTAVRTAFPSGAFPAATBEVKGAKLGFALTDAEFLI
IAEKGLPAKQMPBEVEVLQKASMGATRDNLVYGASNGIDGAIILTPMPYMHATGKI
TONNVLPMTTILARLNLDSQGISVAKEAYETVQLDASKLAEKKKADGKEIKAM
TPGCGHDTMIRWMLAAGIDPDKDVSTIVPPROMVAMKVNMDPVCVGPENOL
VNOGIGFTACTTGELMKGHPERKLGRAADVEKNPNTALLMAVMAQOCDEBANK
EENSTILGRQMFNPFPKOVGLRKNVINGRVLNENTGLQKFWODHASYPFHSHD
SWPTENIRMGKAPDPTDVAKALVAKVRBDIWRMAAKDLGVADLPASTSRGKETPFG
KYFDPENPSAYLESLIKAA"
/gene="AGR_L_1888"
/locus_tag="6067.6866"
/feature="NITRATE TRANSPORT PERMEASE PROTEIN NRTB"
/feature_start=1
/feature_end=1
/feature_note="NITRATE TRANSPORT PERMEASE PROTEIN NRTB"
/feature_id="AAK89515.1"
/db_xref="GI:15159391"
/translation="MPMSALARKFNEEPOOPAVOKANVVALSPKHLPRIDLRKMGAA
AANTPLPLVAVLGVOLCSAPGSSLPSPQVEESYDLIVSPFPNAGSDIGL
LRLVLSERVAVAGFGLAAGIIGAVIGOSVAMRGLDIPQVLRTPVPLAMPLSL
AARODNSPAIFVIFITSTIMPVINTAVGRNIPODYRVAVATLRNOLPEFKIMLP
SAAPYFTGRIGIGISMLAIYAAEMLTGSGVIGFTMDAMNSRLPDIIVLAYIGV
VPMPLDKLVALLGKVTIRGAAN"
/gene="AGR_L_1890"
/locus_tag="6980.7777"
/feature="NITRATE TRANSPORT PROTEIN NASD"
/feature_start=1
/feature_end=1
/feature_note="NITRATE TRANSPORT PROTEIN NASD"
/feature_id="AAK89516.1"
/db_xref="GI:15159392"
/translation="MNAIYKLHDIDYFPKGSRAEVLKGINLTTEKGFVSVIGHSG
CKSTMTNLIAGLTVSAGAVLENNREVEPGERAVVPQNSLFWLSVYNNVILAV
AKVFRNTKTRERHDMVMDLVOVAHAKDKKPSIDSGMQRQVIAIAPALMEPKIL
LIDPEGLDALTRAHLODAVMEIHARLGNVTVMITHDGEAVLISDRIVMTNGPAA
HIGEVLDVPLAIPRNRIELASDKIYKCEAVLKLFEYHRFEAAE"
/gene="AGR_L_1891"
/locus_tag="7714.10230"
/feature="NITRATE TRANSPORT PROTEIN NASD"
/feature_start=1
/feature_end=1
/feature_note="NITRATE TRANSPORT PROTEIN NASD"
/feature_id="AAK89517.1"
/db_xref="GI:15159393"
/translation="MPRGCAEYPLRTSPLRSGVITMTOKLVIIGNGMAPRMENL
FEATGLVDVTIENAEPRVNDRIIMSPVLSGEKSEVEDIVINDENYVANNYTHKGA
KTRIDRANKITSEKGTASTYDRIVATGSLPFIIPVGHOLPGALARDLDVTKM
LTIARKGSAIVIGAGLLEAAYGKRGMDVTVIHLPTMEHQLODPAALVLEKKA
LNERGIDITTKANTKICILGEBKVEGIELDGRILKIDMVAVAVGIRPASGLAKEGIA
VNRGIVDDMMTSDASIALGECABHRCGCVLAPLESARVADLRICGSAEYHG
SVNTKLTGKTVGLNPSADGFAEGDREBIVLRDATGAVYKRLIKENRILIGAVIET
AGSWPFDLMKSTSDISAMRETLIRGOAVQGSPLDPMAVAVALPDNAICGNCYCK
GKITSATISKELTSLDVAHRTKASACGCTGLVEQLTTLTGDTYNAQVQWCK
TDLGHDDVRLIKAKGLKTIIPVMOLEKRTSGCAKCRPALNYLVCDPPEVADY
OSRYINERVHANIQKDGTVSVPRMNGVTNADBLRALDVVDKELPLVKTGQORI
DLGIEKEDLPVAMDIGKAFGISQAVAKGRTVTCVGEOMCRGFTDGLGRI
EKEMGSMTPAKLKLAVSGCPNCACTKQDGIIVDVGSEHIFHFGAAGLKGREY
LGLVHTEDEALHIVALTOMVBOARYLERIYKMAKRIQYDVRQIMDDAKRKAIF
DPEVPSQKFAQVDPMSERVSGHDKHEFRMAAIGSEAAE"
/gene="AGR_L_1892"
/locus_tag="10227.10571"
/feature="NITRATE TRANSPORT PROTEIN NASD"
/feature_start=1
/feature_end=1
/feature_note="NITRATE TRANSPORT PROTEIN NASD"
/feature_id="AAK89518.1"
/db_xref="GI:15159394"
/translation="MRTMSNVNIDIGDISIDPLRGAARCYKTPMGKIAVPTAENQFA
IDHCPHKGEPISQGIVHGASVTCPLHMNVISLETGKALGADEGEVRTIPVRNLDRL
SLALRTLMIAAE"

```

## ORIGIN

```

Query Match
Best Local Similarity 50.2%; Score 133; DB 1; Length 10676;
Matches 572; Conservative 0; Mismatches 495; Indels 72; Gaps 7;

QY 87 CGATGTCGGGCGAGTTGGCACTGAGAAAGAAAGATTTGAATTCGCTTCATCAAGCT 146
DB 4768 CGCTTTCCGGCGCAACCGGAGAACCGGAAGTGAAGGTGCAAAACTCGTTTCATCGCGCT 4827
QY 147 CACCGACATGCGCGCGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 206
DB 4828 CACCGACGCGCGCGCGCTTCATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4887
QY 207 GTTCGTCGAACGTGAAGCG---CAGGCGCACTGGAAGGTGATGATAGGATGCTGA 263
DB 4888 GCCTGAGTCGAGGTTCTGAACAGGCTCTCGGGGGGCGACGCGCGCAATCTGCTGCT 4947
QY 264 TGGCGAATCGACGCGCTCGACATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT 323
DB 4948 CGGCGG-----CGCTGGAAGGCGATGATGCGCGCGCAATCTGATCGCGCGCT 5001
QY 324 CGGCGCAAGCGCGATATCGAGTGGCGCTTCAGCATGCGCTTCGAAGCGCGATTC 383
DB 5002 TCTCATGCTACCGGCAAGGTACCGAGAACAGCGCGCGCGCGCGCGCGCGCGCT 5061
QY 384 GGTGTCGAATGAATCTGCGCATGATGAAGCGCGCAATCGCGCGCGCGCGCGCT 443
DB 5062 ACGCTTAATCTCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5115
QY 444 GGTGATCGCATCAAGCGCGATTAATCTGAACCGCGCTCGCAAAATCAAGCGCGAAG 503
DB 5116 CGGCGTCGACGCTCGATGCTCCAGCTGAAGCGCGCTTCGAGAAAGAAAGCGCGA 5175
QY 504 CAGCGCTTCAATATGCGATGACCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCT 563
DB 5176 CAGGAAATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5235
QY 564 CTGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
DB 5236 TTGGCTGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5269
QY 624 CCAAAATCGCGCGAGCGCTTGTGCGTGAACCGCGCGCGCGCGCGCGCGCGCGCGCT 683
DB 5270 -----GAGTGTCCACCATGCTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5316
QY 684 GGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 743
DB 5317 GAAAGTCGCGCAATGACGCTTCTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCT 5376
QY 744 CAAAGGCGTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 803
DB 5377 TCAGGCGCATCGCGCTTACCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCT 5436
QY 804 AGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 863
DB 5437 GCGCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5496
QY 864 CAGCGCGCTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 923
DB 5497 GATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5556
QY 924 AGCGATCGAAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 983

```

[illegible]

/gene="nirD"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=1  
 /product="nitrite reductase small subunit"  
 /protein\_id="AA144709.1"  
 /db\_xref="GI:17742340"  
 /translation="MDSNWIDIGLIDSDIPLRGARCVTKPMGKIAVFTAEQVAFED  
 HCPHKGSLSGVIGVSGAVSYTCRHNWVSISETGKALGADGEVRTTIPVRMLDGLSIA  
 LETMLIAR"  
 complement(479. .2926)  
 /gene="nirB"  
 /note="synonym: Atu3902"  
 complement(479. .2926)  
 /gene="nirB"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=1  
 /product="nitrite reductase large subunit"  
 /protein\_id="AA144710.1"  
 /db\_xref="GI:17742341"  
 /translation="MTQKIVTIGNGMAPGRMLLENLEFAPGLYDVNTFNAEPNNVOR  
 IMLSVLSEGSYEDYIYHNDERYAANNVTLHKAKATRIDRANKITYTSENGTASID  
 RLVAITGSLPLTIPVPGHQLPGLVATYEDLDVDTMLSLIAGKGAIVITGAGLLGEVVA  
 YGLKROGDVPIVHLMPTIMERQDPAPAAVLLLEKALNERGIDITTKANTCIIIGEEVVA  
 EGILEEDRIVIKGDMVMAVAGIRPASGLAKEAGIAGVAGIVDDGMNTSDASITALDE  
 CAEHRGVMDARVPLYESAVIADRILCGSGAEYHGSVNTKLVKTGINTLFSAGFAGG  
 DDBREIVLRDRTAGVYKRLLIKENRIIGAVYGETDAGSMFDMKSKTDSIMARETL  
 IFGOAYQGGSLDPMMAVAALPDPAELCGCNGYCKKITSATISKGLTSLDVPRAHRT  
 ASASGNGTCELEQLEMLTLTGDTYNPAVAPVPMCKCTDLGHDVNRLLTKAGLITPLV  
 MOELSEWTKSCCAGCRPALNVYLCDFPDEYADRYOSRYINERVHATIQDGYSVVP  
 RMWGGTNAEDLRAIADVDFEELPLVKTVGGQRIDLLGLEKEELPAVMADLGAGFI  
 SGOAVAKLRIKVTQCVGEOMCRFGTOSTGIGRIEKMFGMSWTPALKLAVSGCPEN  
 CAATXCHDIGVTCVDSGEIHPGAAGLDIKGTGVVGLVHTDEDALEHIALQMWYE  
 QARTIERLYRKAKRIGCYDEVYRQRLMDADRKAIYFDFVYSQKFAQVDPMSERYSGHD  
 KHEFPMALIGFSSDAE"  
 complement(2932. .3729)  
 /gene="nirC"  
 /note="synonym: Atu3903"  
 complement(2932. .3729)  
 /gene="nirC"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ABC transporter, nucleotide binding/ATPase  
 protein [nitrate]"  
 /protein\_id="AA144711.1"  
 /db\_xref="GI:17742342"  
 /translation="MNAVYALKLDHIDIKYFDKGGSGRAEVLKGINLTIEKGFVSIVGSHG  
 CGKSMNLNLIGLTRVSAGAVLTLENREVNBPGRPAVVPNNHSLPMLSYVENVNLAV  
 AKVPRNTTRTERERDDWNRNLDLYQMAADCKRSELSGCKMKORVGIRARLAAPEKTL  
 LIDPEFGALDITRAHLQDAVMEIHARLGNTWMIITHVDVBAVLSDRIWMTNTPRPA  
 HIGELVDPLARPNRIELASDKIYLICREAVLYKFLERRHFVEAAE"  
 complement(3743. .4636)  
 /gene="nirB"  
 /note="synonym: Atu3904"  
 complement(3743. .4636)  
 /gene="nirB"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ABC transporter, membrane spanning protein  
 [nitrate]"  
 /protein\_id="AA144712.1"  
 /db\_xref="GI:17742343"  
 /translation="MSALARKFENPEOPAVOKANVVALSPKHLPRIDLRKMGAAAR  
 NTIPPLVLANVLGKWLDCGAPGSSSLPPSPQVREESYDLYVSPFNGSODIGLGR  
 VLVSERAVYFIIAGIILGAVIQSQSWAKMGDPITQVAKTVPLEAFMLPLSLAA  
 FQDPSERAVYFIETSIPIVNTAVTAVNRNIPQYRVNAVAVLRINQLEFFPKMLPSA

```

gene
APY1FTGLRIGIGLSMLAIVAAEMLTGVCIGFPIWDANMSRLPDIIVALAVIGVWG
FMLDKLVAALGKXITGGAAN"
complement (4750. .6042)
/gene="nrta"
/note="synonym: Atu3905"
complement (4750. .6042)
/gene="nrta"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, substrate binding protein
(nitrate)"
/protein_id="AAL44713.1"
/db_xref="GI:17742344"
/translacion="MKKIFSGTVSRRTILKTTATYALVTAVRTAPSGAFAATAEPEV
KAGKLGFIALTDAAPLIIAEKGLFAKHGPEVEVLKQSWGATRNLYVGASNGID
GAHLLTPMYLMTGKVTQNNVPVMTIILRLNDQSGISAKVEATGOLDASLTK
AAPEKKADGKEIKAAATPFGGTHDLMIIRYLAAGGIDDDKDVSTIVPPQVAMK
VGNMDEVCEGPENBOLVNOGIGFTACTTGELMKGHPKALGSRADWVEKRNPAATL
LMAVMAQOCDENANKERKSTILGRQNPVPPKDVGLKGNINYGKRVLENTGL
QMKFWQDHASVPPHSHDSWFTENIRWGFAPDQVLAQVAKVNRDIIWRAAKDGV
ADLPASTSGKEFFPDGVFDPENPSAYLESISIKAS"
complement (6423. .7718)
/gene="nrta"
/note="synonym: Atu3906"
complement (6423. .7718)
/gene="nrta"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ABC transporter, nucleotide binding/ATPase
protein (nitrate)"
/protein_id="AAL44714.1"
/db_xref="GI:17742345"
/translacion="MAPEKTAGSKGATOMRDAPALVGSROKILRAGFIPLVNASV
LTAARSGPADRGILTLVQVSNANVDRALAFODLAHMLSPVPSMGLGSP
PTLTPSPSGRGNATITLSTRLPARKALTGSETTGALENBALVLDIDRGEA
PPLGNTYPPSSHNTYFRYWLVAAGIHPHVDVLLVPPPTSDAALAGAIDGFCGA
PNNTVAERGVRIYAKQDLWPSAPEKIVGKRPWASQOETVGRLLTALDPAASWC
DLANDHDLGALADPRYIGAPQSIIRRYLAGESFISDSQNRVIEKTYTFPGDHANY
PROSGLTIYSQIMRWGAELSETGVNALSAYRPDIYRALGDKAPEDADIRIEQ
DSDRPFVDSGFDPADIAGVNVSFAVRISVPVSABSGEA"
complement (7754. .8341)
/gene="nastr"
/note="synonym: Atu3907"
complement (7754. .8341)
/gene="nastr"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="two component response regulator"
/protein_id="AAL44715.1"
/db_xref="GI:17742346"
/translacion="MTLRDLTILVIDENAIRASITIEGREGYQRTVYIHEVNGAR
TLETLOPDIYFIDLENPNDMMEHLFQLRTVGRPIAMVEDSDASIEAAVEAGVSA
VYVDGLKKEKVRPILDMAVSRFNASRLQRELADAKSALEBRKVIERAGILMKRGL
SEEEAFALRQANMEKKISIEIAQSVTAAGLGM"
complement (8443. .9138)
/gene="Atu3908"
complement (8443. .9138)
/gene="Atu3908"
complement (8443. .9138)
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AAL44716.1"
/db_xref="GI:17742347"
/translacion="MLHCSRSRSMATEIEGLPLADRLAKGIRTLISGELAPGRLS
EAFAFSRPFVSNLSREAFRLTLKQGLLHEANRGVFAVVPMTMSVIDIYRRARVNEC

```

```

Query Match 9.6%; Score 133; DB 1; Length 10707;
Best Local Similarity 50.2%; Pred. No. 1.9e-15;
Matches 572; Conservative 0; Mismatches 495; Indels 72; Gaps 7;
GALROAMHHPAIIKMARAVENDAKTIRREABDLVGVSAAMKFEHAAVVDLODSARISDF
YRIIAELRLCTGLKDPQQLSPFDVMSKILSLMEBKAGDAENMELYANLSERT
QY 87 CGATGTCGGGGCAATGTCGAGTGTGAAAGAAAGATTGAAATTTGGCTTCATCAAGCT 146
DB 5941 CGCTTTCGGGGCCACCGGCAAGAACCGGAAATGAAAGGTGCAAACTCGTTTCATCGCGCT 5882
QY 147 CACCGCATGGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 206
DB 5881 CACCGACGCGCGCGCGCGCTATCATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5822
QY 207 GTTCGTGCAACTGGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263
DB 5821 GCCTGAGGTTCGAGGTTCTGAAACAGGCTCTCTGGGGGGCGAGCGCGCAATCTCGTGT 5762
QY 264 TGGCGAATGACAGCGCTCGCAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 323
DB 5761 CGCGCGCGCGCGCGCTCGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 5708
QY 324 CGGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 383
DB 5707 TCTCATGCAATACCGGCAAGGTGACGAGAACAGCGCGCGCGCGCGCGCGCGCGCGCG 5648
QY 384 GGTGTCATGCAATGCAATGTCGATCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
DB 5647 AGCGCTTAATCTCGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5594
QY 444 GGTGCAATCGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 503
DB 5593 CGCGCTCAAGCTGATGCTTCCAGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5534
QY 504 CAAGCGCTCAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 563
DB 5533 CAAGGAATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5474
QY 564 CTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
DB 5473 TTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5440
QY 624 CCAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 683
DB 5439 GGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5393
QY 684 GGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 743
DB 5392 GAAAGTCGCAATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5333
QY 744 CAAGGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 803
DB 5332 TGAAGGCAATCGCGCTTTATCTGCTGTACACCGCGGAACTCTGGAAGGCGCATCCGGA 5273
QY 804 AGTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863
DB 5272 GCGCGTCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5213
QY 864 CAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 923
DB 5212 GATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5153
QY 924 AGCGCATGCAATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 983
DB 5152 CACGATC-----CTCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5099
QY 984 TAGCATGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1040
DB 5098 CGGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5039
QY 1041 CAACGCTTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1100

```





TAHKALDASASABWAAWAKERGLSVGTNGCPDILHAGHVSILHAASOCRLVIG  
LNSDASVRLKGGCRPVNDQHPRACTLALASDANVVEEPRLIALILPDLIVK  
GADYITDVTGADVQKAGRVVLDLVAGKSTGTIGLAKRGSTS"  
/complement (6734. .7366)  
/gene="m112564"  
/complement (6734. .7366)  
/gene="m112564"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB49663.1"  
/db\_xref="GI:14023056"  
/translation="MPSTELLAFPATTAIPAYIPGPAMLYAAOITLARGMSGJLTA  
LGHIGGVHVPAAAGSLVPLHAPVPLMAYKGLALILGSLPRKRGVVAL  
PALERSARAPFESITVEYLNKPTAIFPMALPQDLSALPWLQFVGLTVNL  
MFSADIVCPFLGLVLSRLRRSRAQRMVQAGAVLVLGVALQRS"  
7499. .8449  
/gene="m1r2565"  
7499. .8449  
/gene="m1r2565"  
/codon\_start=1  
/transl\_table=11  
/product="ADP-L-glycero-D-mannohexose-6-epimerase"  
/protein\_id="BAB49664.1"  
/db\_xref="GI:14023057"  
/translation="MTIYGCAGMTGSNTVAALNMGHDLVVDLTGHTANLAD  
LSIADYLDQDFLTRIAGALGLIEAVFHQACSTTENNKGFMEEVNAVYKRLHA  
CAAPRVPLVASSASVYGGSEEREPLERPLANVAYSKLFDVRYRTVDTJSH  
VAGLFPYNYGPREAHKGMASVAPHLFNQOVERGPNPLFGADGPGQSRDFLHV  
GVADVNLMLMKRSGSIFNCGTGAQPRALAEVYIDTLGKEIEIFPFDLKGSY  
QSFQADMRLRLAAGVNGOFRITETGVRYVEMLAKQRS"  
8422. .9429  
/gene="m1r2566"  
8422. .9429  
/gene="m1r2566"  
/codon\_start=1  
/transl\_table=11  
/product="ADP-heptose-LPS heptosyltransferase"  
/protein\_id="BAB49665.1"  
/db\_xref="GI:14023058"  
/translation="MAESPILLIVIGPRWDMWMAQCFLSALKELYPNNAIDVLA  
APVLRKMEPIRQIDPLPKPALERTIRRGRLRGVDAVYLPGWKGSLALPEF  
ALIPRFGHLREMYGLTDIYPLDPAVRTOAVFSLAKGSRAPPLATYDANQDA  
ALLDRGLAPRFRFALMGAEERGPAPKWSSEYAGIKRPMKGLKVALFGKNDQV  
TNETLAPGVVDLQOTRLEDAIDILAAILAVSNDGLMVAAGVGPVAVVGT  
SPENTPPLAERRELVWLHLSLSPCHQKVCPLGLHNCLEKTLFVGQVATAERLLEPAT  
A"  
9426. .10400  
/gene="m1r2567"  
9426. .10400  
/gene="m1r2567"  
/note="unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB49666.1"  
/db\_xref="GI:14023059"  
/translation="MKVLIVKTSMSGDVHTTPAVEDSLARPDISPCKVEERAGI  
VALHPAIGIHTVAIRRMKALFNGTWEBAALRRTRRECFDVIQAQGLKALV  
AQAARAPIDPSSARSBSATLFYDVYGVPRNLAIERTRLFGALGYQPDLSL  
ASGIYPPGALAGTGTAFLLHGSREDKKVPEDMIGTARLIVERGMTPTTWSNE  
REKSVAAIARAVPVVVKSPADIALLGRSLVIGADTGLTHLSAFLPPIYAV  
PLATPGLTGPFGYASTILAVPGRLTPAEVVAERDLSLNLKQAPA"  
complement (10535. .11008)  
/gene="m112568"  
complement (10535. .11008)  
/gene="m112568"  
/codon\_start=1

Query Match 8.8%; Score 122.8; DB 1; Length 348077;  
Best Local Similarity 51.7%; Pred. No. 1.7e-13;  
Matches 394; Conservative 0; Mismatches 317; Indels 51; Gaps 3;  
QY 455 TCAAGGCAATATATCAAAACCGGTGCTGGAATAATACAAAGCCGAGCAAGCGCTTCA 514

DB 227011 TCGATGCCCTCCAGACTGAAGACCGCTTGCCGCCAAGAGGACCGGCAAGAGCTCA 227070  
QY 515 ATATGGCGATGACCTTCCCGCGCGATCTCACAACATCAAACTGCTTACTGCTGCGG 574  
DB 227071 AAGCGCGCATGACCTTCCCGGCGGAGCAGATGACCTTGGGCTGCTACTGCTGCGG 227130  
QY 575 CTGGCGGTATCAATCTGCTGCTACTATTCGCCCGCGCAGACATTTCCGCGCAATCGCG 634  
DB 227131 CCGGCGGCAATGATCCCGCAAG----- 227153  
QY 635 CAGAGCGCTTGTGTGCGGACCCCGCGCGGAAATGCGGTCACGCTGGAAGCGCGCA 694  
DB 227154 --GAGCTTCGACCATGCTGCTGCGCGCGCGCGCAGATGCTGCGCAATGAGTGGCA 227211  
QY 695 CCATTTTCGGTATATGCGTGGCGAGCCGTGAAACGAGAGCCGCTTCAAGGCGATAG 754  
DB 227212 ACATGAGCGCTTTCGCGTGGCGAGCGGTGAAAGAGCAGCTGTCACAGGCGTGC 227271  
QY 755 GCGTCCGCGTATCACCGATGAAGAACTTGAAGACACGCCGAAAAAGTCTTGGCG 814  
DB 227272 GCTTACCGCGCGCACACGCGGCGAATGTGAAAGGCTCATCCGAAAAAGCGCTGCTGC 227331  
QY 815 TCACCAAACAATGGGCGGAGAAATACCCCAACCTATCTGCGGAGACCAAGCGCTGA 874  
DB 227332 TCGCGCGCGAATTCATCGAAGAAACCGACGCCCAAGCGATCTGATGCGCTCA 227391  
QY 875 TTCGGCGCGCATCTGCTGCGAGCGCGCAATACAGAAACCGCAAGAGCCATCGAA 934  
DB 227392 TCGAAGCCAGAGTGTGCGGCGCAAGAGAAACAGAGATGCGCGCATCATCG 227451  
QY 935 TCGTGGCGCAAAACATATCGTGGTCTGACGTGAAGTCTGCGCGCTAGCATGAACG 994  
DB 227452 GCAAGCGCAATGATGAACGTCGCGGTGCGCAATCATTCGCGCTTAAAGGCGACA 227511  
QY 995 GCACCTTGAATACGAAAAAGACGATTAACGGCGCTACCGGACTTCAACACTTCTTC 1054  
DB 227512 TCACTACGCGAATGCGCGCGCGCGCGCAAGGCAAC--CGATTTCTACATGAAGTTCTGGA 227568  
QY 1055 GCCACGCGCGCAGCTATCGTCTCTACAGAGTGAAGTGTATGTGACCCAGCTGAGCG 1114  
DB 227569 AAGGCGGCGGTTCTTCAACCTTCAAGAGCAGATGATGCTGCTCGCGGAGAACATCG 227628  
QY 1115 GCTGGCGCATGATCAATGAATCAAAACCGCACTGCTGATTCGATACCGCGCAAGACG 1174  
DB 227629 GCTGGGCGCA-----AATTCGCGCGCGCACGACGCTCAAGCGCGTGTGACAGG 227679  
QY 1175 TCTACCGCGGACATCTATCTGCGCGCGCGCAAGAACTGG 1216  
DB 227680 TCAACCGCGAGATCTGTGCGCGAGGCGCGCAAGAGATCTCG 227721

RESULT 10  
AP005945  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE

AP005945 300150 bp DNA linear BCT 12-SRP-2003  
Bradyrhizobium japonicum USDA 110 DNA, complete genome, section  
11/31.  
AP005945 BA000040  
AP005945.1 GI:27350985  
Bradyrhizobium japonicum USDA 110  
Bradyrhizobium japonicum USDA 110  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobiales; Bradyrhizobium.  
1  
Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uehi, T.,  
Saenmoto, S., Watanabe, A., Iidesawa, K., Itiguchi, M., Kawashima, K.,  
Kohara, M., Watsunoto, M., Shimpo, S., Tsunoka, H., Wada, T., Yamada, M.  
and Tabata, S.  
Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110  
DNA Res. 9 (6), 189-197 (2002)  
22484998

PUBMED 12597275  
2  
AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T.,  
Sasamoto, S., Watanabe, A., Idegawa, K., Iritani, M., Kawashima, K.,  
Kohara, M., Matsumoto, M., Shimo, S., Tsunoka, H., Wada, T., Yamada, M.  
and Tabata, S.  
TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110 (supplement)  
JOURNAL DNA Res 9 (6), 225-256 (2002)  
MEDLINE 22485002  
PUBMED 12597279  
3 (bases 1 to 300150)  
AUTHORS Kaneko, T.  
JOURNAL Direct Submission  
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail:kaneko@kazusa.or.jp,  
URL:http://www.kazusa.or.jp/rhizobase/,  
Tel:81-438-52-3935 (ex.2338) / Fax:81-438-52-3934)  
FEATURES  
source Location/Qualifiers  
1..300150  
/organism="Bradyrhizobium japonicum USDA 110"  
/mol\_type="genomic DNA"  
/strain="USDA110"  
/db\_xref="taxon:224911"  
/gene="b112729"  
complement(14..862)  
/note="ORF ID:b112729  
probable ABC transporter permease protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47994.1"  
/db\_xref="GI:27350986"  
/translation="MDNRNERMTPTROIIGKIGLMSFVIVSPALFELMNLISLKL  
FEIDNAAYPPVFEPEHIAWKNYADVLASNRFLTYFINSLVTSATLALIVGPAGY  
GIARMAAHKSAIVILARITPGISYLIPLFLFQMLGLGLTVPOIILHVVTVSIVI  
WMIGVFETTPLELEBALIDGATRWQFRHVALPAKGIAPFLIAVIFESMNFVF  
GIVLAGRETRTIPVAVYTMISFDQSLMGPLAALIVLPULLTVFAQRQIVAGLTA  
GAVKGG"  
complement(835..1773)  
/gene="b112730"  
complement(835..1773)  
/note="ORF ID:b112730  
probable ABC transporter permease protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47995.1"  
/db\_xref="GI:27350987"  
/translation="NSALTQADPAAARSDAPEKELRPSYFPVVVPALVVLITITF  
PWFPTIWMSLNWKVGSPTTEVGLSNLYRLTSDPRTIEVGHITVYTALSVLPLVLG  
TAAAVFHQKFAGRGLRGVIMPMAATPVALAWTMHFLQGLVNLISLVGIPA  
QLVVHPATVPLSLVETWQMTPLVMLVLGLALAIPTPEESAOIDANIMQMFRE  
ITLPLIMPLFIAGMIRMTDAVKSPFIIPAIIOGGGSGSETINITYLSVARTYVDLG  
VGSATVAVPELLIVLAAVMLYARQRMWVTEIASDA"  
complement(1770..3152)  
/gene="b112731"  
complement(1770..3152)  
/note="ORF ID:b112731  
probable ABC transporter substrate-binding protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47996.1"  
/db\_xref="GI:27350988"  
/translation="MARSLVSLPKTMSHREDPEVVDYRISRTLLTGTAAAGALSITG  
LPABAENVMKYAGTKLEVIILKGRGDMLCKNIKEFTELNGIOYEAEOIPROOROK  
VITFETSGRPSFDVYHLSTHYOKROPEKAGMLADMTPEPKDLSLTAPDLVESDFSAIG  
LDYAKNDKGQMLSPWSVDYFTLYNKELFQKRGVAVPRTLDENVAAAEKLTDPKEGT  
YGFVGRGLRNAMMALMTNPFNLVYGGFILDCKNIIITLDGEALTAATKLYOTLTITKSAFP

gene  
CDS  
complement(3179..4012)  
/gene="b112732"  
complement(3179..4012)  
/gene="b112732"  
/note="ORF ID:b112732  
putative cytochrome C"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47997.1"  
/db\_xref="GI:27350989"  
/translation="MDRRQRLFYHRLGIDHGRMKFWRALAVATLLAAAPALLAGEIAP  
BARRGVAFMGPBDRAMQDDDTANBGMFLYVDGELMTKTKGSAAKACADCGDARG  
MKGVAAYPAREKAVGRPTLIDQRLNLCRAHQQTPLPYESRDLLALSAFAHDSRG  
VAITGADDPQKQFVQGRDLMQREGLNACINCHDNDPDRLAGAPITQAPFTGY  
PLYRLWQTLGSLERRLRSCMSGVAAQYDYGAPBLVALLEFLYMSRARGLPWETPAVR  
P"  
complement(3955..4272)  
/gene="b112733"  
complement(3955..4272)  
/gene="b112733"  
/note="ORF ID:b112733  
probable sulfur oxidation protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47998.1"  
/db\_xref="GI:27350990"  
/translation="MAALINVPAKARQDVIEIRLTQSHINETGRHRTADGRLVPRDI  
ITSFTCRVNGTEVFRAIDLFPALIANPYLSFTVAKESGKFEFMTGNDGYSSTASASI  
TVE"  
complement(4272..4739)  
/gene="b112734"  
complement(4272..4739)  
/gene="b112734"  
/note="ORF ID:b112734  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC47999.1"  
/db\_xref="GI:27350991"  
/translation="MPTTRQFLSLAGITTAAGTIPYTLRPLEATPAMLTAINRVV  
GEAPIRTGVLKDIPLVENGNTVMTSVASPMTPDDYVSGSHVENKNPQNTGNF  
HLGPWAGSAQVSTRIRLADTKVVALIATSDSFTQVADVVYTLTAETSEEN"  
complement(5385..6665)  
/gene="b112735"  
complement(5385..6665)  
/note="ORF ID:b112735"  
/codon\_start=1  
/transl\_table=11  
/product="flavocytochrome C flavoprotein subunit"  
/protein\_id="BAC48000.1"  
/db\_xref="GI:27350992"  
/translation="RRPNAPVTRRNVTGITAATSYAVPSIILBAOSAGRVVYVGG  
FGGAACALTRANKAOVTLIEPKVTVISCPFSNEVYAGIRETEAODRGVDKIYABG  
ITVVAQAATTVDPERSVTTTDDGATLTYDLVLSGIDFPYBALFGYDEASERKPHA  
WKGAQTLRLRQLEAMEGDTVAIAIPANSRCPAPYERASLTIAHLTKTKPKPAKY  
LILDAKQSFQOURLFERAMELYGMIMIERIGLSQGRVTSVASTKTLVTEGNTPTD  
VANVIPPORAGALQIAGADATGWCPIDPVTFESKLVPNIHVIGDACICGGGI PKGAS  
AASQGRCAAGAIIVILAGRAPETPRLTGVCYTTVA PGYGSIAQNYQPRGDIYAWE  
GGATSPVDAPREIAREADEAARMQGITTAATPG"  
complement(6782..8968)  
/gene="b112736"  
complement(6782..8968)  
/note="ORF ID:b112736  
putative aldehyde dehydrogenase protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC48001.1"

```

/db xref="GI:27350993"
/translation="NMKHVSPKMNRRAPVIGTAAGAGLIGLIDIPROGPAVVRADG
SEBICAMVYVRPDDTVIRIARISEMGQSLTGLAQVABELECDMSKSTERYPTGQS
VAKRVMGDFSTGSGRIGRSQDYRKGGATRVMLIQAAISVMPVASECTAASVIT
THTPSGRITTTGKVAEAAKLTPEPADVTKDPRWMLJGKVTLRIDTADKTGTVYG
IDVKLPGMLNAAIKDCPVFGKLSYDEAKIAGMKGVKVVGTAVAVVADTWMHA
KPALEALPIVMEDEMAKVSASIAKMLEGLDNAPAVGKNDAKAKAIAAGVAKV
EAVVYVYONNHAETMEPMNATLYTAEKCEVWCGTONGEAPAAITLEASGLPAEKDVH
KMPGGGFRRCQDTYVROAVIAKOMPTPIKLMSREDAHGRHPTTCCKATGA
PADNNVALHRLSGQSLFSLRPEALONGDPAPAFQVAGSGLAAGFYSVNLLIE
HAMRNPVHPGFWRGVNVNHAIMECFMDELAAAGODPLEFRRLKMSHDKTAVL
NAVAKIGWTTAPQGVYRIGIAQWGVSGYVAGAAISVTDGNKTKVHRIVASTPGY
VYNPAVERIOIAGSFYQGSALPYGCGCTKDGKIDQNTBDTYSNRIEMPVESIMV
PSGFMGSGEPTTGYAAPAVINAPAAIKRIRSVPLBDQNTTA"
/gene
/gene="b112737"
/complement(6980..9456)
/gene="b112737"
/contig="b112737"
/contig_start=1
/codon_start=1
/translation="MDMTNGTAMNLTJNGSLTLDVDPDPLMAIRENGLTGT
VCGGIAOCGACTVHIDGATRAQGVSAABEKKITTEGLASGSLHRYOAMTAKD
VFCGTCQSGMTMAVVAALINEKPKFDADIDALITNICKGTFQVREAIHTTASA"
/gene="b112738"
/complement(9678..10070)
/gene="b112738"
/contig="b112738"
/contig_start=1
/codon_start=1
/translation="ORF ID: b112738
unknown protein"
CDS
gene
8.6%; Score 119.8; DB 1; Length 300150;
Best Local Similarity 52.4%; Pred. No. 6.4e-13;
Matches 404; Conservative 0; Mismatches 307; Indels 60; Gaps 4;
Query Match
470 TAAACCGGTCGTCGAAAAATCAAGCCGAAGCCGCTTCAATATAGCGATGACCT 529
80688 TGAAGCGCGCTTCGAGAGAAAGAGCCGAAGCCGCAAGTCGCGATGACCT 80747
QY
530 TCCCGCGCGGATCTCAACAATCAAGCTGCTTCTGCGCGCTGCGGATATCAATC 589
80748 TCCCGCGCGGATCTCAACAATCAAGCTGCTTCTGCGCGCTGCGGATATCAATC 80807
QY
590 CTGGCTACTATTGCGCGCGCGAGACATTTCCGCGCAATCGCGCGAGCGCTTGTGT 649
80808 CTGACAG-----GACGTCGAGACCA 80828
QY
650 CGGTACCCCGCGCGCAATGCGCTCCACGCTGGAAGCCGCAACATTTTCGTTATT 709
80829 TCACGCTGCGCGCGCGAGATGCTGCAATGAAGTCGCGCAACATGAGCGCTTTC 80888
QY
710 GCGTCGCGAGCGCTGGAACAAGCAGCGCGCTGTTCAAGGCAATAGCGCTGCGGTATCA 769
80889 GCGTCGCTAGCGCTGCGCGCGAGCTGCTCCACGAGGCAATCGGCTTCAAGCGCTGCA 80948
QY
770 CGGATGAAGACTCTGGAAGAGCAGCGCGGAAAAAGTCTTTCGCGCGTGAACCAATAGG 829
80949 AACCAGCGAGATCTGAGACCAAGCATCCGAAAAAGTCACTTCGCAATGCGCGCGCTGAG 81008
QY
830 CGGAGAAATATCCCAACATCTATCTGCGCGTGAACCAAGCGCTGATTCGCGCGCATCT 889
81009 TCGACAGAAATCGAAGGCGCGCAAGCGCATCTCTGATGCGAATAGAGAGCCAGCAT 81068
QY
890 GCGTCGAGCGCGCAATAACAGAACCGCAAGAGACCAATCGAATGCTGCGCGCAAAAC 949
81069 GG-----GCCGACAAGATGAGAGAACAGAGAGCTCCGCAACCATCATGAGCGCAAGCGCG 81122
QY
950 AATAGCTGCTGTGACGTGGAAGTGTCTGCGCGCTAGCATGAACGCACTTTCGAATAG 1009

```

```

DB
81123 ACTGATATCACTGCCCGGTCGAGACATCGCGCATTCGCCCAAGGCAAGTTCGACTAG 81182
QY
1010 AAAAAGCAGTAAACGCG-----CGCTACCGGACTTCAACACTTTCTTCCCAAGCGCG 1063
DB
81183 GCAAGCGCGCGCAAGGTGTCGAGAACTCGCCGCAATCATGAAGTAACTGCGCGCATCAG 81242
QY
1064 CAGCTATCCGCTCTACAGCAATGCAAGTGTGTATCTGACCCAGCTGAGCGCGCTGGGGA 1123
DB
81243 CCTCTATCCGTTCCAGAGCCATGATCTCTGTTGATGACCGAGACATCCGCTGGGGA 81302
QY
1124 TGAATCAATGATTAATCAACCGGCAACTGATATCTGATATCCGCAAGAGCTCTACCGCG 1183
DB
81303 -----AATACGAGCGCGCTTCGACAGCAAGCGCTGATCCCGCAAGTTCACCGCG 81353
QY
1184 CGGACATCTATCTGCGCGCGCGCAAGAACTGTCGCGAGAGCGCAAGCGCA 1234
DB
81354 AGGACCTCTGAGAGAGCGCGCGCAAGCGCTTGATGTCGCAAGCTTCGCA 81404

RESULT 11
RME591985_07/c
WPCOMMENT
Sequence split into 17 fragments LOCUS RME591985 Accession AL591985
Fragment Name Begin End
RME591985_00 1 110000
RME591985_01 100001 210000
RME591985_02 200001 310000
RME591985_03 300001 410000
RME591985_04 400001 510000
RME591985_05 500001 610000
RME591985_06 600001 710000
RME591985_07 700001 810000
RME591985_08 800001 910000
RME591985_09 900001 1010000
RME591985_10 1000001 1110000
RME591985_11 1100001 1210000
RME591985_12 1200001 1310000
RME591985_13 1300001 1410000
RME591985_14 1400001 1510000
RME591985_15 1500001 1610000
RME591985_16 1600001 1683333
Continuation (8 of 17) of RME591985 from base 700001 (AL591985 Sinothizobium melioli 10)

Query Match
8.5%; Score 118.2; DB 1; Length 110000;
Best Local Similarity 48.7%; Pred. No. 1.3e-12;
Matches 487; Conservative 0; Mismatches 463; Indels 51; Gaps 4;
QY
132 CGGCTTCAATCAAGCTCACCGAGATAGCGCGCGCTGCGCGCGCGGCAAAAGCTTCTT 191
DB
56467 CGGATTCATCCCTCTGCGAGTGCAGCGCGCTGATTCGCGCGCGCGGAAATTCGCGCTTGC 56408
QY
192 CGAGACGAGGCGCTGTTCTGTCGAACGAGGCGCAACTGAAAGTGTATGGA 251
DB
56407 GCAAGCGCGAGGCAATCACCTCTGAACTGTCAAGACGCTCTCTGCGGAAATGTGCGCA 56348
QY
252 TAGGCTCGTAATGCGAATGAGACGCGCTCGCAATCTGCGCGCGCGCGCTTAGCGCG 311
DB
56347 CAGGCTCGCTTCTGCTAGTTGAGCGTGGCCACATGCTTCAACCATGCGCGGTGCGCG 56288
QY
312 CAGCGTTGCTTCCGCGCAACAA-----GCGGATATCGAGTGCCTTTCAGATGGCTTCA 368
DB
56287 CATGCTCGGCTTCCGCTCCAAATCCGTCGCGCGAGATTCGCGCGCTTTCCTGAGGCGCG 56228
QY
369 CGGCAACCGATTAAGTGTCCAAATGAATGATGATGATGATGATGATGATGATGATGATGAT 428
DB
56227 CGGCAACCGATTAAGTGTCTGCAACGCGCTCTAGAGGCTGATGATGATGATGATGATGATGAT 56168
QY
429 GGAAGCGGTAAACCGGTGCTATCGATGATGATGATGATGATGATGATGATGATGATGATGAT 488
DB
56167 CGGTGAGGCGGAGATGCTGTAAGAGCCCAAGGCGCGCTGCGCGCGCTGATTCGAGCGCG 56108
QY
489 AATCAAGCGGAGGCAAGCGCTTCAATATGCGATGATGATGATGATGATGATGATGATGATGAT 548

```

```

Db      56107 GGGGGAGCGCGAAGCCCGCGCTGACGCTCGCGCGTACACTTCTTCTGCGCATTA 56048
        |||||
Qy      549 CATCAACTGCGTTACTTGCTGGCGGCTGGCGGATTCATCTTGCTACTATTGCGCGC 608
        |||||
Db      56047 CTACGAGTTCCGCTACTGCTGCTCGCGCGCGCATTCGATTC----- 56006
        |||||
Qy      609 GGAAGACATTTCCGCGCAATGCGGCGACAGCGCTTGTTGTGGTGAACCCCGCGCGCA 668
        |||||
Db      56005 -----CGACCGCAAGCTCAAGCTGCTGTGTGTGCGCGCGCAAT 55967
        |||||
Qy      669 AATGCCGTCACGCTGGAAGCCGCGCACATTTTGGTTATTGCGTGGCGAGCCGTGGA 728
        |||||
Db      55966 GACCTCGAAGCGCGCGCGCGCGCGCGATCGACGCGCTTTTGGTGTGCGCGCGCGGA 55907
        |||||
Qy      729 CCAAGACGCGCGTGTAAAGGCAATAGCGCTGCGGATTCACCGATGAAGAACTTGGA 788
        |||||
Db      55906 CATGTGTGATCCGAACGCGCGCTCGCGCATGTCGCGACCAAGCAGACATCTGCGC 55847
        |||||
Qy      789 GGAACACCGCGAAGAAAGTCTGCGGCGTGAACAAAGATGGGCGGAATAACCCCAAC 848
        |||||
Db      55846 CTGCGACCGAAGAAAGTCACTGCGATGCGTCCCGATGAGCGAGGCTAATAGGATAC 55787
        |||||
Qy      849 CTATCTGCGCGTGAACCAAGCGCTGATTGCGCGCGGATCTGCTGACCGACATTA 908
        |||||
Db      55786 GGTCTGCGCGCTGTGTGTGCGCGCTGACCGCGCGCGCTGTGTGGAACCGGCGCA 55727
        |||||
Qy      909 CAAGAACCGCAAGAAAGCATGAAATGCTGGCGCAAAACAAATAGCTGTGCTGACGT 968
        |||||
Db      55726 CCGCGCGCGCTGTG-----CGAAGTCTGTGCAAGAGCGTCACTGCGCAGCGCGCG 55673
        |||||
Qy      969 GGAAGTCTGCGCGCTAGCATGAACGCACTTTCGAATCGAAGAAAGCATTAACGCGC 1028
        |||||
Db      55672 CCAATCATTCGCGCGGCTACTTGGCGCAATTCGCAATCGACCCGAGGAAACCGCGC 55613
        |||||
Qy      1029 GCTACCGGACTTCAACACCTTTCTTTCGCGACGCG---GCCAGTATCTGCTTACAC 1085
        |||||
Db      55612 CGTCGCGAAGAACTATCTGCTTTCACGAGGCTTTGGCAACTACCGCGTCCGACGA 55553
        |||||
Qy      1086 TGCAGTCTGTATCTGACCCAGCTGAGCGCGCTGGGCGCATGA 1126
        |||||
Db      55552 GCGGCTGTGACCTACAGCAGATGCTGCGCGCGCGCAGA 55512
        |||||

RESULT 12
AL646058/c 196050 bp DNA linear BCT 02-SEP-2002
DEFINITION Ralstonia solanacearum GM11000 chromosome, complete sequence;
ACCESSION AL646058 AL646052
VERSION AL646058.1 GI:17427202
KEYWORDS
SOURCE Ralstonia solanacearum
ORGANISM Ralstonia solanacearum
REFERENCE 1
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Laveie,M., Moisan,A., Robert,C., Saurin,W., Schex,T.,
Sigler,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Nature 415 (6871), 497-502 (2002)
MEDLINE 11823852
PUBMED 2 (bases 1 to 196050)
REFERENCE Boucher,C.A.
AUTHORS Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,

```

```

FEATURES
source
    /organism="Ralstonia solanacearum"
    /mol_type="genomic DNA"
    /strain="GM11000"
    /db_xref="taxon:305"
    117..1154
    /gene="RSC0194"
    /note="synonym: RS00626"
    117..1154
    /gene="RSC0194"
    /function="small molecule metabolism"
    /note="Product confidence : probable
    Gene name confidence : hypothetical
    predicted by Codon usage
    predicted by Homology
    predicted by Framed"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="PROBABLE ZINC-DEPENDENT ALCOHOL DEHYDROGENASE
    OXIDOREDUCTASE PROTEIN"
    /protein_id="CAD13722.1"
    /db_xref="GI:17427203"
    /db_xref="GOA:O8Y2Y7"
    /db_xref="UniProt/TREMBL:O8Y2Y7"
    /translation="MKALVYEGPKISLENPKRELQAPGAVVAVTLTTTCGTDLHI
    IKGDPNCEPRTIGHEGVGESVATSKPDHYLISCSGCGANRCGMYS
    HCEHGWTLGHRIDGQVQAEYRVPAEISLTKVPQVDEALVMSDFPCEFGS
    NGRVPGSTVAIVAGPIGLASLTLAQTSPARITIMITNVAKLEVASGLGSDCIN
    TREMAAANVKRLTDMGVDCALIAVGPAPELQCELVAPGVANIGHGVKADLH
    ETLMSONAIATRLVDTVSTPMLTKTVAGKLLPQQLITTHRFKLDIILAAYDTFRNAG
    ETHALKVIIT"
    1164..3017
    /gene="RSC0195"
    /note="synonym: RS00627"
    1164..3017
    /gene="RSC0195"
    /function="miscellaneous; hypothetical/global homology"
    /note="Product confidence : hypothetical
    Gene name confidence : hypothetical
    predicted by Codon usage
    predicted by Homology
    predicted by Framed"
    /codon_start=1
    /evidence=not_experimental
    /transl_table=11
    /product="CONSERVED HYPOTHETICAL PROTEIN"
    /protein_id="CAD13723.1"
    /db_xref="GI:17427204"
    /db_xref="GOA:O8Y2Y6"
    /db_xref="UniProt/TREMBL:O8Y2Y6"
    /translation="MRRRLRGTPCMGMQAGLTPRELSPVCAVLAADRYLAGETTP
    EQVRRVVAHAAAEAPDRTRHPARLPVYMKLHGAICGRIMANAGVDRQATMNCV
    HPVIRPGQVAGAADIDALKQAOVTLMGSGVGVDFSAVPMDLQGEODGTSVC
    EVIRBFDTCALALANTRGQMAVLCNHPDLAFAAARGKRWTFPNSVAVTD
    AEMQAVADDAQMRQLRAEPASQRAAGAHLLDGMHCYATVPARQMDMLVEAHS
    AEPGLFTDITNANDLALETIATINCGPPLPAGSCVLPINLSMDIOPPGVGS
    GQMPEDFPVSPGAVHIOVRMDNALIDIRWMLIEMREAHAKRRIGCVGTGADALT
    MGLPYQAPBARNLAVQIRCIKDNNAYASASALAQGRSPYLPFRASRLAPGFVATLP
    GAVERTTARHGLRNSHLSTAPGTSVSLAFGNCSSGIEPAPDAVYQRRVRIHQCPQ
    LVRIENAHYRFOAMGHEHAVLPSTVFANOVDDGDIKMYAALQPFIDASISKTLII
    SGVSPAVQVALLFRWQLRLKGLITIFRPDPSDBEVLVLPADGRPVCFC"

```

```

gene
3343. .4683
/gene="RSC0196"
/note="synonym: RS00628"
3343. .4683
/gene="RSC0196"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/translation_table=11
/product="PUTATIVE CYTOCHROME C1 SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD13724.1"
/db_xref="GI:17427205"
/db_xref="GOA:O8Y2Y5"
/db_xref="UniProt/TREMBL:O8Y2Y5"
/translation="MSHLIALKAFWLLIGLLANSGEAGAVPAFARQTGACMACH
VSEPLTPGRGFEKMTYLSNSTIPVSGMVSTNRTTIDODFEDVRNGDTVPQ
QASLEFYAGRIVGVEGAFQMTYDGIHSHALDNTDIRAAMHLREDDVDFYGLTVNN
PTVSDVMSSPAQGYRYAANIGVPLAGLIDGIAQOVGSAVAFORHYVLAIG
AYRTADASVPAFGODTATPGVABLAGAPYRAVYNGEASHLMFTGGLIANG
YEDNTMPTGTDRFDYATDAQYQYITMAHAFQAQAMTIEHODMPSATSPSGIGAP
TPANPTDLNAEFKASYYQRYGATLAVFTTGADPLYAQAAPVTSANGRPQTQ
GVIFELDYLPHPKIALQYTWYKENGANVYDGNAINNTIYLLGMWFMF"
4719. .5339
/gene="RSC0197"
/note="synonym: RS00629"
4719. .5339
/gene="RSC0197"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/translation_table=11
/product="PUTATIVE CYTOCHROME C SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD13725.1"
/db_xref="GI:17427206"
/db_xref="GOA:O8Y2Y4"
/db_xref="UniProt/TREMBL:O8Y2Y4"
/translation="MKGFHQPRRVMLAGVITLATTGMSVTCVAGNDEKLAQLCASCH
GYHGSSEPMFRDLQMPHYLEADLKGRDRGRTDQAQFMWGIASQUDVTIASL
ABYVARTQAPAGADPALTARGKEIFEGILPAEGPAPACSHGQAQNDTFPRLAG
OHEAYLRLQISYFKNGTRANAPVMSAVHTLTDEQAKVAFAFLQAO"
complement(5374. .6015)
/gene="RSC0198"
/note="synonym: RS00630"
complement(5374. .6015)
/gene="RSC0198"
/function="miscellaneous; unknown"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage"
/codon_start=1
/evidence=not_experimental
/translation_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD13726.1"
/db_xref="GI:17427207"
/db_xref="UniProt/TREMBL:O8Y2Y3"
/translation="VYNLGLFVFLVNLVIAFAPPTMWNVSWIGTOPANPALLA
GIAACASTGRVLARLSCTLRNMRMRADRONIDVIRTMGEGRKMTGVLVYAF
SPPSNVYIAYGLTGMRLMLIGPEFVGVRWSTYLTWNTIAQISFARYDESEINGAY
MSIYFVLSQMLFALVYGPAKVDHNIITKRLMRRRRPPAPSSDDPRATND"
complement(6044. .6568)
gene

```

```

CDS
/gene="RSC0199"
/note="synonym: RS00631"
complement(6044. .6568)
/gene="RSC0199"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Homology"
/codon_start=1
/evidence=not_experimental
/translation_table=11
/product="PROBABLE CYTOCHROME B561 CYTOCHROME
TRANSMEMBRANE PROTEIN"
/protein_id="CAD13727.1"
/db_xref="GI:17427208"
/db_xref="GOA:O8Y2Y2"
/db_xref="UniProt/TREMBL:O8Y2Y2"
/translation="NRFEKYSLSGRALHMLVFLAALIAVVTIELDAPPKSARAA
LFAAQITGLSVLALMLVRLVRVGQVPAPVSGQLERRARLTHGLVYLMGVMPI
LGVLAIVMSGRKIQPFGLATLTLLADKTLNLAKVHESGATLVYIFVGLIAAALW
HEFIIKDLRLRRMI"
7807. .8229
/gene="RSC0200"
/note="synonym: RS00632"
7807. .8229
/gene="RSC0200"

```

```

Query Match      8.3%; Score 116; DB 1; Length 196050;
Best Local Similarity 54.5%; Pred. No. 3,4e-12;
Matches 305; Conservative 0; Mismatches 240; Indels 15; Gaps 3;

```

```

QY 637 GAGCCTTGTTCGTGATGACCCGCGCGCAATATCCGTTCACGCTGAAACCGGACCC 696
    |||||
DB 195304 GACGCCAAGGCGATCACCGGTGCGCGCGCGAGATGTCGCAACATGCGCGTGGCAAC 195245
    |||||

QY 697 ATTTTCGTTATTGCGTCGCGCAAGCGTGGAAACGACGAGCCGTGTTCAAGGCGATAGCC 756
    |||||
DB 195244 ATGGAACGGCTACTGCGTGGCGCAAGCGCGCGCGCGCCATGCGCCACGCGATCGGC 195185
    |||||

QY 757 GTGCGCGTATCACCGATGAAGAACCTCTGGAAAGACAGCGCGGAAAAAGTCTTCGCGCGT 816
    |||||
DB 195184 TTTCACCGCGAGACACACGAGCGAGATCTGAAAGAACATCTCCGAAAGGTGCTCGGACCC 195125
    |||||

QY 817 ACCAAACATATGCGCGAGAAATATACCCCAACCTATCTGCGCGTGAACCAAGCGCTGATT 876
    |||||
DB 195124 ACGGCGAGTTTCGCGCAAGATATCGAACACCGCGCGCGCGCTGACCGCGCCGCTGCTG 195065
    |||||

QY 877 CGGCGCGGATTTGCTGGAACCGCAATTAACAAGAACCGCAAGAAAGCCATCGAATG 936
    |||||
DB 195064 GAGGCGCTCCAAATTCATGATCTCCGCGCTCGAACCGCGCGCAACAGCGC--CGAGACG 195008
    |||||

QY 937 CTGGCGCAAAAACAATACGTGCGTGTCTGACGCGGAAGTCTGCGGCGCTAGCAATGAACGCG 996
    |||||
DB 195007 GTGGCGCCCAAGTCTTACGTTCAACACCGACATGACATCATCTGACCGCATGCTCGCG 194948
    |||||

QY 997 ACTTTCGAATACGAAAAAGACGATTAACGCGCGCTACCGGACTTCGAACACTTCTTTTGC 1056
    |||||
DB 194947 CGCTACACCAACGCGCTGCGCAAGACCTGGGACGATGGGACCGGATGGCGTTTACAC 194888
    |||||

QY 1057 CAGGCGC---CAGCTATTCGTTCTTACAGAGTGAAGTCTGTGATCTGACCCAGCTGAG 1113
    |||||
DB 194887 GACGCGCGCGTCAACTTCCCGTACCTGTGCGACGCGCATGTGTTCTTACACAGCAG 194828
    |||||

QY 1114 CGCTGGGCGATGATCAATGAATTCAAAACCGGACACTGTGATCTGGAATACCGCGCAAGAC 1173
    |||||
DB 194827 CGCTGGGCGCTG-----CTGAAGACGACCCCGACTACCTGCGCATGCGCAGGAG 194777
    |||||

QY 1174 GTCTACCGCGCGGACATCTA 1193
    |||||
DB 194776 GTCAACCGCGCTGACATCTA 194757
    |||||

```

RESULT 13  
LOCUS AP005951/c  
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 17/31.  
ACCESSION AP005951 BA000040  
VERSION AP005951.1 GI:27352606  
KEYWORDS  
SOURCE  
ORGANISM Bradyrhizobium japonicum USDA 110  
Bradyrhizobium japonicum USDA 110  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Bradyrhizobium.  
REFERENCE  
AUTHORS 1  
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriuguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110  
JOURNAL DNA Res. 9 (6), 189-197 (2002)  
MEDLINE 22484998  
PUBMED 12597275  
REFERENCE 2  
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,  
Sasamoto,S., Watanabe,A., Idesawa,K., Iriuguchi,M., Kawashima,K.,  
Kohara,M., Matsumoto,M., Shimpo,S., Tsuruoka,H., Wada,T., Yamada,M.  
and Tabata,S.  
Complete genomic sequence of nitrogen-fixing symbiotic bacterium  
Bradyrhizobium japonicum USDA110 (supplement)  
JOURNAL DNA Res. 9 (6), 225-256 (2002)  
MEDLINE 22485002  
PUBMED 12597279  
REFERENCE 3 (bases 1 to 298700)  
Kaneko,T.  
Direct Submission  
Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 252-0812, Japan  
(E-mail:kaneko@kazusa.or.jp,  
URL:http://www.kazusa.or.jp/rhizobase/  
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)  
Location/Qualifiers  
1. 298700  
/organism="Bradyrhizobium japonicum USDA 110"  
/mol\_type="genomic DNA"  
/strain="USDA110"  
/db\_xref="taxon:224911"  
48..512  
/gene="b1r4344"  
48..512  
/gene="b1r4344"  
/note="ORF\_ID:b1r4344  
unknown protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAC49609.1"  
/db\_xref="GI:27352607"  
/translation="MTQATPASPSTPKRRPRQVWLTDAAQRISLTQRADSETVG  
LKVGVNCGCAGCSTVEYAHDRPFDVEVDGVKITLDPKAVLFLSHRGDGLGRQ  
DAGPVRLQQPQPLRLRLRIGRAAGGEDRRVGPALPFIHSAVLAARHTTA"  
584..1069  
/gene="b1r4345"  
584..1069  
/gene="b1r4345"  
/note="ORF\_ID:b1r4345  
hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAC49610.1"  
/db\_xref="GI:27352608"  
/translation="MGPGFRQDDDERGAMDBEFLIDLFADPGPVTTIRMGSGYGISA  
DGTFLALRAGLFLRADELTIIDFETEGSKPFGVSTRAKTYTVNSYEWLPAFLPDSS  
OETLAQWRAALAAORAKVKORPKGKSKKVAPEKKVPARKRPAAKCAVRRVQAKRAH

gene  
CDS  
R"  
complement(1134..1664)  
/gene="fabA"  
complement(1134..1664)  
/gene="fabA"  
/EC\_number="4.2.1.60"  
/note="ORF\_ID:b114346"  
/codon\_start=1  
/transl\_table=1  
/product="3-hydroxydecanoyl-[acyl-carrier-protein]  
dehydratase"  
/protein\_id="BAC49611.1"  
/db\_xref="GI:27352609"  
/translation="MPNPDHPPTQPSYTRBELSSSEGYPGNGAOLPAPPMILMD  
RLSEISLDGSRGKHIVGELDIYGHWFEPFCHPRGDIPLBSGLDMMQWIGWIG  
MSGSPGKRRAIGVGEVCTGETTTPVQVRIVAMRMVRGRGLVGIADGRVLADGAC  
VFTADMKVGLTQAVE"  
complement(1763..2830)  
/gene="b114347"  
complement(1763..2830)  
/note="ORF\_ID:b114347  
hypothetical protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="BAC49612.1"  
/db\_xref="GI:27352610"  
/translation="MKVLDHERHMAFAEVALGQIRSLRQTAIPRNYEIWYVATGY  
NAPLKTINETLARGKLTLEADLEQIYETIYSHIKTTIRIDVNGARVIGEIDVVKV  
GDALGMTSGSYDASSGATERLSAKSRQIKAVITLVRSISEMRETKALIEDRLTIS  
RNEISNLCQSSLEAIRABESLTDPLTGLGNRYKFDYRMIGNAVOSALASGPLSLLEPDI  
HFKSPNDYGHLTGQDVRLVGLSMKQTIKQODIPARYGGEFAVLIPNTALRQALTY  
ADHIRAVAMKEIKKSTGTEILGRTTISVGSMLKPGDDTALIERADACLYAAARG  
RNRVCEADPREPAVETHSRVA"  
complement(2994..4496)  
/gene="rhlE"  
complement(2994..4496)  
/note="ORF\_ID:b114348"  
/codon\_start=1  
/transl\_table=1  
/product="Atp-dependent RNA helicase"  
/protein\_id="BAC49613.1"  
/db\_xref="GI:27352611"  
/translation="MSFENLGLSEKVLAAVATGYTPTPIQEOAIPIHYLARKDVLGI  
AQTGNGKAAFTVLPMLTLEKGRABRBPRTLLPTEELAAOVKENVDRVAGQKIN  
VALIIGVSPEDQDAKLTGRGVULIATGRGLDHTERGLLITGVLELYIDBDMKD  
MGFLPDIERICKLVEPTRQTLFTATWPEIRIRITETRELANPQKVESKPAITAVTV  
OSOVPAGRKAHEKRELRLREAQDLKNAIIFCNRKREVAIVHSLQHGFSVGAH  
GDMQDPARMAALEQFRKGEPLVLVASDVARGLDIPEVSHVNPVPVPHIPDDYHARG  
RTGRARGSTAIISIVTPLDOKSWAIEKLIGQSIIPRAGDVEVHAAGDATTAPREOR  
GRERSRGQRGRQGRDRDRSHEPREARPSAEARPARPARPREPREAREPSREPR  
HGSRQANASHVPSIGRPEPRRQREADTEPDGSHLPALFAPVNSPAGA"  
complement(4787..5782)  
/gene="b114349"  
complement(4787..5782)  
/note="rnc family  
ORF\_ID:b114349"  
/codon\_start=1  
/transl\_table=1  
/product="transcriptional regulatory protein"  
/protein\_id="BAC49614.1"  
/db\_xref="GI:27352612"  
/translation="MTRMCTDVEVEQEPFYWRBYRANGLSGVTALELERADPFGE  
FSLRQLAGAGLVELKASPYTVERSTSDIAYTPGDSICVYQDQSGGWFGKMGKSPAI  
LIACSPATSHTDLPYRTAPDAGFRLRLIKIIVPSRIPOADRIRILTPTFENDHAP  
IIAAGCFADIGVAAAGDGDGAADATSLVQTSIVQALADRLIERGIVRGSLAQ  
AALRTARRLSOARRLIARQLONPDLAPNAVADLIGVSHLHLFPAEKSFSQOTYDE  
RLKRSRLRLRAPEHLINDIATSCGFESLATTYYRVFNAAYGMAPDFARVVEGH"  
5935..7503  
/gene="b1r4350"  
5935..7503



```

/gene="b1r4350"
/notes="ORF ID:b1r4350
hypoetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49615.1"
/db_xref="GI:27352613"
/translation="MDCTGAGGKSPSCNGNFSNTRRLFRILVTLGLAALGFPALPAC
AERVALVGNNDYKRVFKLKAIVNDARTMDTLKQFGSVVAENQSQSQSETLLA
PDALIPGDIAPFFVAGHGEIAGQNYLLPTDVPATQEQBELVRDASLIADRIVERL
ONKARTSLIVPACNPNPEERKTRAVAGGGLAEMQOLPEGVFSFAGSPROTLDN
RLSNDADPNVSFTFTRFAKELQGENLVQVORTRLVSEMAUTVXHOVAPYPMOM
VDVFLSGAKDAARADBPPOKIALAPVSVPRVPEKETTNAPIASFRRNGASV
VFSRADPTIGISWRMANGDFRETGFITDILDRTRKMRNPSIELEPPDAQATIEVRY
VDOSDMQGFPIKRDPEALIRDKRIIDMTATSLSFREENGILLVYTHLVSYRCA
IEVRIGIDTAVPNQVLPKMPGCDMDPSAITAGMPLMKLAPATQFVSVELTYRDSV
SEIKSFRANRNN"
7987.9162
/gene="b1r4351"
7987.9162
/gene="b1r4351"
/notes="ORF ID:b1r4351"
/codon_start=1
/transl_table=11
/product="PAD-dependent monooxygenase"
/protein_id="BAC49616.1"
/db_xref="GI:27352614"
/translation="NMLGYLKGAGIDVVLVLEKADPFEDFRGDTVHPSTLOVDELG
LIDGFLKPHORLOKMDLFGGTPVRIADLSRLNTKYPIAEMPQMDFNLFREGKR
PSLRVMTTEADVLIRGETFAGVAKPEGGIDIEADLTACDRSHTVBERGKR
VVEIGAPMDLVPRAGRKADTESVFAVPEKETTNAPIASFRRNGASV
KARGLOALDDVVRAPPIRSGLADYKSDIDILTVAINRLARTRPGLITGDHAY
ANSPVGAVNALVQDAVATANLADKLPHGCSBEDLAVRRREFPRKMQAOVI
VQNNIISGALQGDRLKXVPLIVRLITLALPMQIGPARIALGVREHVSQAAPAS"
9644.10108
/gene="b1r4352"
9644.10108
/gene="b1r4352"
/notes="ORF ID:b1r4352
hypoetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC49617.1"
/db_xref="GI:27352615"
/translation="MFGMGITISGLDNKGSFLNTAFGADGVFSRANLATVGRAG
FAVQNNLFYIKGIVAGVNNRLSVSDVGGATSGGQTHMANGMTVAGAGMEYVTRMI
VGLVYVYAFSGQTYQLCGTTNYTFDAKPRDLQMAVVRASVYKFPAPVIARY"
complement(10205.10975)
/gene="kdu"
complement(10205.10975)
/notes="kdu"
/notes="ORF ID:b114353"
/codon_start=1
/transl_table=11
/product="2-deoxy-D-glucose 3-dehydrogenase"
/protein_id="BAC49618.1"

Query Match 8.1%; Score 113.2; DB 1; Length 298700;
Best Local Similarity 54.9%; Pred. No. 1.2e-11;
Matches 267; Conservative 0; Mismatches 213; Indels 6; Gaps 2;

QY 131 TCGGCTTCATCAAGCTCACCGACATGCGCGCGTGGCGGTCGCCGCAAAAAGCTTCT 190
DB 263952 TCGGGTTCTATTTCGGCTGTCATGCGCGCGCGTGTATCTTCGCGCGTCAAGAGGCTTTA 263893
QY 191 TCGAGGACGAGGCGCTGTTGTCGTCACACTGGAAGCGCAACTGGAAGGTGTGATG 250
DB 263892 CCGCGCCCAAGGCTTCGACGTCGAACTGCTGTGCGGCAAGGTCTTCTGTCAACGTCGCG 263833
QY 251 ATAGGCTGTGAATGCGAACTGACGCGCTCGCAATGCTGGGCGCGCGCTTAGCGG 310
DB 263832 ACAAGCTCAATATCGGCTGTTGACGCGCGCATATGCTGTGGCGCGCGTGGCGATGCGCT 263773

```

```

QY 311 CCAGCGTTGGCTTCGCG---CACCAAGCCGATATGAGGTGCGCTTACAGATGGCTTCA 367
DB 263772 CTTCACTCGGCGCTCGGCAAGTCAAGTCCGATCCGCGCTTCAATCTCGGATCA 263713
QY 368 ACAGCAACCGGATTAACGGTGTCCAAATGAATCTGCGATCAGATGAAGCGCAATACCCG 427
DB 263712 ACCGCAACCGGATCAGCGTGTGCGGCAAGCTTCATGCCGCGCTGATGAGAGATGAG 263653
QY 428 TCGAAGCGGTTAAACCGGTGATCCGATCAGCAGATTAATTAACCGGTGTGAA 487
DB 263652 GCGACCGCTTCATCGGCTCGTTAGCGGCAAGGCGCTGCTC---GGTGTGCGCAGG 263596
QY 488 AATCAACCGCAAGGCAAGCGCTTCAATATGCGCATGACTTCCGGCGGATCTCA 547
DB 263595 GCGCAAGCTAGTGCATCCGATCCGATCCTTGCGGATGACCTTTCGTTCTGACCCACA 263536
QY 548 ACATCAACTGCGTTACTGCTGCGCGGCTGCGGATCAATCTGCGTCTATTCGCGCG 607
DB 263535 ATTACCAATTGCGGCTTCTGATGCGGCGCGCGCGCTTATCCACAGATGTGCGG 263476
QY 608 CGCAGG 613
DB 263475 TGGTGG 263470

RESULT 14
AE007223/c
LOCUS AE007223 12129 bp DNA linear BCT 03-JUN-2004
DEFINITION Sinorhizobium meliloti 1021 plasmid pSyma section 29 of 121 of the
complete plasmid sequence.
ACCESSION AE007223 AE004669
VERSION AE007223.1 GI:14523386
KEYWORDS Sinorhizobium meliloti 1021
SOURCE Sinorhizobium meliloti 1021
ORGANISM Sinorhizobium meliloti 1021
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
1 (bases 1 to 12129)
REFERENCE
Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Hutzar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kalan,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Pederspiel,N.A. and Long,S.R.
Nucleotide sequence and predicted functions of the entire
Sinorhizobium meliloti pSyma megaplasmid
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9883-9888 (2001)
JOURNAL
MEDLINE
PUBMED 21396509
11481432
TITLE 2 (bases 1 to 12129)
AUTHORS Barnett,M.J., Fisher,R.F., Jones,T., Komp,C., Abola,A.P.,
Barloy-Hubler,F., Bowser,L., Capela,D., Galibert,F., Gouzy,J.,
Gurjal,M., Hong,A., Hutzar,L., Hyman,R.W., Kahn,D., Kahn,M.L.,
Kalan,S., Keating,D.H., Palm,C., Peck,M.C., Surzycki,R.,
Wells,D.H., Yeh,K.-C., Davis,R.W., Pederspiel,N.A. and Long,S.R.
Direct Submission
Submitted (29-MAR-2001) Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA
FEATURES
source
1..12129
/organism="Sinorhizobium meliloti 1021"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:266834"
/plasmid="pSyma"
/complement(103..318)
/locus_tag="SMA0580"
/complement(103..318)
/locus_tag="SMA0580"
/function="Miscellaneous, Unknown"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypoetical protein"

```

```
/protein_id="AAK64962.1"
/db_xref="GI:14523387"
/translation="MSSRALQSTVWGIGQFVWAKAFRARSDDSSRTACPDAAIKG
ALVGATKPEPRSRQTSALQADGFPVNG"
complement(394..1194)
/locus_tag="SMA0581"
complement(394..1194)
/locus_tag="SMA0581"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction Excellent match (e-87) to
nitrate transport protein ATP binding component nasd from
Klebsiella oxytoca starting at amino acid 35 of glimmer
prediction. Met33 of glimmer prediction has good SD
sequence and was called SMA0581 start. Chrd, Nasd and Nrtc
are synonyms. Possible operon with SMA0581, SMA0583 and
SMA0585"
/codon_start=1
/transl_table=11
/product="Nitrate transport ATP binding protein, probable"
/protein_id="AAK64963.1"
/db_xref="GI:14523388"
/translation="MTKSYLSLELDKSPFGRTREYVKQSYLVVDKGFISIGHS
GGKSTLVIVGSLDTATGTVVLDGKVIDEGPDRAVVFQNSLPLVTEENRLA
VKVFSRTNKQERHEWTNRNELVQMAHAABKBEVSGMKQVGIARALAMPEKV
LILDEPFGALDALTRAHLQDQVQIHATLGNVIMITHVDVAVLISDRIVMTNGPS
ARVGEILDVPLARPRRIELASDRYLTCTRESVKLFYERHREVAE"
complement(1208..2101)
/locus_tag="SMA0583"
complement(1208..2101)
/locus_tag="SMA0583"
complement(1208..2101)
/locus_tag="SMA0583"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction Possible operon with SMA0581,
SMA0583 and SMA0585. Excellent match to nitrate transport
proteins in Klebsiella (Nasb) and cyanobacteria (Ntrb)"
/codon_start=1
/transl_table=11
/product="Ntrb, Nitrate transport permease protein"
/protein_id="AAK64964.1"
/db_xref="GI:14523389"
/translation="MSGVTLKLRKPAQTPOTAQVIALGOTASRGIDGRSPVQTV
NILPLVLTFTPTLMAQILCSSPESLPSPRYLBSRWELIHPRTIGGVDGILFMH
VFASLQVALGYMAAAVAGLGTGQSLMRGDLPIFQVLRVPPPLAMPLSLAA
PDGTPGSAIFVIFITAIWPIIINTAVGINIQQDYQNVAKVRLNSFEYFGIMDPA
AVIFGRLGIGLSWLAIVAAEMLIGVGIGFPIWDANSSLSIDIYALIVGIVG
FLDLRLIALVGAVTRGTANA"
complement(2164..3456)
/locus_tag="SMA0585"
complement(2164..3456)
/locus_tag="SMA0585"
complement(2164..3456)
/locus_tag="SMA0585"
/function="Cell processes; Transport of small molecules"
/notes="glimmer prediction Excellent match (e-143) with
nitrate binding protein precursor proteins from several
cyanobacteria. Possible operon with SMA0581, SMA0583 and
SMA0585"
/codon_start=1
/transl_table=11
/product="NrtcA-type periplasmic nitrate transport binding
protein, probable"
/protein_id="AAK64965.1"
/db_xref="GI:14523390"
/translation="MTKTLFGILTRRSVLTKTATAMVGAARTLLPSGAFAGAPET
AAATLGFIALTDSAPLIIAKEKGLFDKYGTEVEVVKQASGCTRDNLVLGAGAGID
GAHLITPMFYLISTKVTONOPLPMAIILARLNLQOASVGAAYADLKVGIDASVKG
DAFAKKKGAAGAUKVAMTFPGGTHDILITRYMLAAGIDDPKQVETIVVPPQMVAMK
VGTMDQFCVGEPRANEQILVNOKITGYTVNTAEIWAEPKESPMARDWTEKPRAYKAL
VWALEAAQWCDMANDEKLAKIVGRSWFNVPKPIDVRLGGEYDYGKIVENSPIH
FKPFMEHSAISYFQSHDAFPLETENRMGLASDTDIKGIATVNRBEDIWREAAKDLGI
SIPASTSRGPETFFDGKVPDPANPETYKSLAISRIA"
4274..4732
/locus_tag="SMA0590"
4274..4732
/locus_tag="SMA0590"
/function="Miscellaneous; Unknown"

repeat_region
gene
CDS
/locus_tag="SMA0591"
4761..4955
/locus_tag="SMA0591"
4761..4955
/locus_tag="SMA0591"
/function="Miscellaneous; Hypothetical/Partial homology"
/notes="glimmer prediction Excellent match to 64 amino
acids at carboxyl terminus of much larger (44 kda) protein
from Sinorhizobium NGR234"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein, fragment"
/protein_id="AAK64967.1"
/db_xref="GI:14523392"
/translation="MVRSRWYRKLNRFRAGIEAGISGLTTRYGLAHCTRGLHHEFT
YVSSVVAYNLALFARLPRT"
5249..6421
/locus_tag="SMA0592"
5249..6421
/locus_tag="SMA0592"
5249..6421
/locus_tag="SMA0592"
/function="Miscellaneous; Hypothetical/Globular homology"
/notes="glimmer prediction match (e-19) to two similar size
(45 kDa) proteins in Sinorhizobium sp. NGR234 over entire
length"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK64968.1"
/db_xref="GI:14523393"
/translation="MALRSANILPDEETAGLVETANGTRFAVHSDNNEGIIACFP
STOREHEKVGILHPFOHLGPGWLRBOQASAHVVERDDIGILIRVGADICGAVSIR
PPDAQQLPRTETATVSGRTVSGVQKLLTYKQDENFVRPASAGSLYTAKEPSDR
IDNLRNLSLNRWTAALVGEREYVGFASITFAVVDETALVTRFDRPNQKRLSD
CAQLLSKRGQDYAGKYDAAYEDIAAIIHQSSRAPIDLLFPNRLIVFTLIGNDAAH
LKNPSLLETPTGLRLSPAYDVVNTAFYDGPOTLALSTGSKIHLEAANOAIFFAFGX
EIGLPDRRIDQTFQQLKQVAKASIIIRPPDAEPADGVHFRKEIVNSCURILET"
6397..7428
/locus_tag="SMA0594"
6397..7428
/locus_tag="SMA0594"
6397..7428
/locus_tag="SMA0594"
/function="Miscellaneous; Unknown"
/notes="glimmer prediction"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK64969.1"
/db_xref="GI:14523394"
/translation="MPENTGNIAPLAWQRYVEBALRRRKAGELTQKHSLAAGVSHPT
MAAFBERGTTTLTKALDILKVLGVNDEPTEGDOARVVRAPFERWRVLVPLPQDSP
ARFNGGWPRPVLVYGLKMLKEGLTAFBEILKEKAYVRLKQGWGPPLPIREAIQPEVDL
LIECHLADQGEVEVRGENDPAHCCFMRABSGRMFLIRGYEDGAKETPAAGTIDTLT
PLMKRGVLLAEKXILSLRDADYAVVHFRAMFTGIRGLVLSWANPLSDLLVEGH
GARDDEATLEKXFSANDIESRLACMPLTLTLYERFGVAGLSINRVAEYORLNSP
ISKREPRPR"
7698..7919
/locus_tag="SMA0596"
7698..7919
/locus_tag="SMA0596"
7698..7919
/locus_tag="SMA0596"
/function="Miscellaneous; Unknown"
```



GATLRVAGAGAGVINTLSETPKGPKLTLKTEAGSYGFSRYVYKGGDDTFSYNLY  
YGLVAFNNFPYSLTIPNTAQFYERDVAENQCSLFGYLPKELGAVTLTSGINDTGN  
ASDSYSGKLVPRDPANRLTLRLNOONSANBAEPGRSAFLGCRGTSVQPGTTSGR  
RLPLDANGNELPCOTQAFIEMTPPTAPEGSRVFNAPNANAGATVAPGOSYLIEPTGCT  
TNFPQRFLOAQTEBATVDFNDYDITPTTSLNSYLSVRLTLTTTPQSPFPDTNARGTCT  
PCALVTPPPAQPPIIDNQRYEAOIALNSQISPGQONSFGINLEDSYQOQSKAQTE  
FPBRLSRSLFLIDISFSPNELKANGLRLTNSQPSGVLTTPAAGVRSPPNNVLSIRA  
NYSYIFNAPALSNLFATGSNILENGLRPGESQVTVGVGDLTPANRLGAPTFYNSYL  
DCAIGGVSLANPNPATAAIPLVLOQONQDTRQASGIEBAVADMPTBOLRLQATWMS  
TVRAGNADNPQSQSTYPPYQODPBGIPENSYLASVYHOGMIVALVNRXGCKRRA  
TTASGVSRGNSNEYPAWATLDLNEIPLVPLFTLGSVFNLTDIYESLSGIPAPGT  
TRIGRLELG3  
complement (3934. . 6291)  
/gene="g111963"  
complement (3934. . 6291)  
/gene="g111963"  
/note="ORF\_ID:g111963  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC89904.1"  
/db\_xref="GI:35212530"  
/translation="MRRRVYLGSLILLVYGAKLAANVATKIATKPYIQPGDADG  
IADQVIAWQTANGPDLYTVDFGKTPYGRVIAGRVNVHLYSSDPLPQPSYASG  
AMNNYALGLEYETRYFCVGRGGLPLEGCASFYRKRSAGSFVWVGDFEPH  
SETOKYELHYOARLIVAMSRLEELKFPTPALPRPDIINTGDNVYIIGSGQRFH  
KMPVNSDYVACDDGAPLARSLLPIYVAINHDIIGAGDIYNLLADADAPNPLPVGE  
RSGALGEGDALAEVNNYVPLANGPGNDPQYIPGDCARAGFTPAQGTKYRBPAA  
IEAYRSTTVPTLSCPRQIDHNSNFSFYAAHFVFDANPHELDARVAGANSADK  
SPPFGSPYSPVLDLWLDLDRSEQLMRFVYVHQPFSATATRYQMKALARVLEDD  
GANLVENGHNHYORTHPLRAGRVAEADTVGHPAVDTAFDGIAGRVPGLIHI  
EGSGGFTNSGKLEDSAVGRSAAAGALPGAGWDTLTCARMAPVANAARGPKI  
TARFKARVPSFGVNVVDGPHLITLYOISEPLSATCSAADPPYSGDADPNKAPLA  
DLPLDPAEGLVTPATIGTPALIDKFTYRPLDGLSYELHVEHPTGTSOPLPORVA  
LVNRSYALANGTOVVLGPTGYTPRAIYRKKEKCTQERSVYTVVGRLEBGAQKTYTIE  
CVVAEAPDPVQALITASVYSTALEPVAKPVTLQSQSARTRAVR"  
complement (6288. . 7748)  
/gene="g111964"  
complement (6288. . 7748)  
/gene="g111964"  
/note="ORF\_ID:g111964"  
/codon\_start=1  
/transl\_table=11  
/product="dimethylamine monooxygenase"  
/protein\_id="BAC89905.1"  
/db\_xref="GI:35212531"  
/translation="MKRCVAYIGGIGISGVAKCLRDGHQVTLYESIDYQGIWY  
RTSGGTFESVRFQNSKYLSAFSDYPMPEQMSDFPHTEILAYLNSYDHFELRCIR  
LNCQVEKRSRDMKVTIVSTPEGASSEFDALACSGVFRPWPNTIPEGADFPGLT  
LHAKDYKEPSMFAKRVVVMNGASGVDAVATDPAQCVISFRANSLVFRYFAGR  
PLDCNLSRLYALIPAGVRKSLYSRKPAPIEAMHROCNLEPAGLFPASIPANEPYDL  
VAKAIETRPITASGSGORVLTDSSTGADLVIYATAGVSPFPDDASVYVNHGCT  
DIYKXVHPDLPRNCGFIGIIRITGALLPFAEMQARFSVYLSOHLPTBESMABEQ  
NRPAQOQKQWVASGRSFOVROVEYTEELIARLIDAPKWRHWOJAMQULLPTPLPTH  
YRLDGNPKWKGSEWIRQVLDIKIPRNASGVGVR"  
8491. . 12084  
/gene="g111965"  
8491. . 12084  
/gene="g111965"  
/note="ORF\_ID:g111965"  
/codon\_start=1  
/transl\_table=11  
/product="WD-repeat protein"  
/protein\_id="BAC89906.1"  
/db\_xref="GI:35212532"  
/translation="MKTFRKQKQGVLLTPQGYSRLOEARRLFRGNFGRNTLDM  
SRTGLAPRTLKWSRKGVDRQTLQCFHTFGLOLNEDEYLSIQIGRGRQPHAK  
PMHELEGAPNFAQFYGRQDELATLKQMIQVRSRLVAVLGMGIGICTICLAKLQIHI  
LEFEVYLRSVNGLPYKELLADLVAFHPOBELPASFEGOVYSLAHAKKGLIV  
VLNDIETWSSGSLAGRLVPRGFEVYGLFEGAVSGTANHCIVLTSCDQDQGVNFECK  
RIPVRLYRLGIGLTHBGRQILLEKGLAABEGPNNKELIRCTAGNPALAKIYASTIRE  
LFDGQANFLQOQSVAFGIVDLEQHYRLSDILECJYMLAVNREPVLLSQLEDY  
LIPVRSRLLEAMESLRARSLIECGGTFTLLPVLEYTTDLLEALIVREIVDGKSVL

FRRLAIKQSPBEHVKAQIRFLKPLVDOLLAAVFSRLALEHYTHRLJGELRSELQ  
OLGAVAGLNLILCHCEILEGYPSTLTVQADLRGLRCLNANVQNFQSITFEA  
LSTVSVASPDQOLATSEINGITRLMQADQADLRCRHTSNVMSIAPSPDRVL  
ASGDAFVRVLRMDYRTGQICLPVCGHREBWSVAEFHPGCGILLASGSEDAVRMEVDS  
GRCLTLTRHSGHGMTHAVRFSFNGOMLASSQDDKIQLMHPSSEGLQMOQHTLASD  
IAFAPDQGLTISGSDQDQTLRLMDVQRGLLKLQGHTEWVSVDFSAGRTLASGSD  
QTVRLMDSDSLCVRVMHGNMWSVVSFSDGRLTSGVDHSVRWEISGCLRV  
LQHGSGGIVSVAFGDGKTLASGSIDHSVRLMDESTROPMKSLQHTSMRVTVASPD  
GTLASGSGODRTITLMDPDSGRCLKTLRSGHGWNSLAFSPNGALLASSVDHSRLV  
NVEHQCCGMIGQHTSWRSVAFHPDGRVLASADQKTARLMDIETGCLMTQOHTS  
WRSVAFHPDQHTLASGSDQDTYKLMVDQTRLADSLSGHSGVSVFPADGKRLAS  
GADQTVRLMDTSMQCTHVLNRHAGVLCAVATENDSILASSSADETITLMDLQGN  
YIGTMRLEGPYTMNITGATLGRASQALTKALGAVERDAAPATPSGILDR"  
complement (12067. . 12756)  
/gene="g111966"  
complement (12067. . 12756)  
/gene="g111966"  
/note="ORF\_ID:g111966  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC89907.1"  
/db\_xref="GI:35212533"  
/translation="MGDRTGLAVALLVTYLAGLPALGSPALAEFLVAVGCSFF  
IDQFLLQRLSLVLVGLIOERRPGRREVRVHBEHGLVHARLGLPVGVTYLGWESF  
RRGQGGGCVTLGSPPALGIEGESYCATYKALAEHMYSEAVGAVEQOQVGL  
LAAQRTGAVVAARLTBRNERRAAALIDAEHQTHNALAERMLTQGLERCQLLNSTK  
DAASDSAVQX"  
complement (12761. . 13045)  
/gene="g811967"  
complement (12761. . 13045)  
/gene="g811967"  
/note="g811967"  
/note="ORF\_ID:g811967  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC89908.1"  
/db\_xref="GI:35212534"  
/translation="MPKATNGTVLAQSDKCEVVEGNYPPPDALQTEYFQPSNTRTT  
CPWGVASVYVVVDSKVNKDAWYVPPHQAHHITQYIAFMKOTRVEA"  
complement (13242. . 14666)  
/gene="g111968"  
complement (13242. . 14666)  
/gene="g111968"  
/note="ORF\_ID:g111968  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAC89909.1"  
Query Match 7.5% Score 104.8; DB 1; Length 299750;  
Best Local Similarity 52.2% Pred. No. 4.9e-10;  
Matches 260; Conservative 0; Mismatches 232; Indels 6; Gaps 1;  
QY 624 CCAATTCGCGCAGCGCTTGTTCGTTGATGACCCCGCGCGCAATGCCGTCACGCT 683  
DB 142467 CCAACCGGACCGGACGATGATCGACCAACGCTGCGCGCGGATGATGCAACCT 142526  
QY 684 GGAAGCCGGCACCATTTTCGTTATTCGTCGCGGACGCGGTGGAACGACGAGCCGTGTT 743  
DB 142537 CAAGCCGGCAACATCGACGCGCTTTTGATCGGCAACCTCTGGAACATCCGCGCGCGC 142586  
QY 744 CAAGGCGATAGGCGCGCGGTGATCAGCGATGAAGAACTCTGGAAGAGACAGCGCGAANA 803  
DB 142587 CGAGGCGATTCGCTTTACGATCGCGACGATCTGGAGATCTGGCCCCGTCACCCGGGCA 142646  
QY 804 AGCTTCGCGGTGACCAACATAGGCGGAGAAATACCCCAACCTATCTTGCGCGTGC 863  
DB 142647 GGTGCTGGGGGCGCGGACGCTGGCGGCAACCTACCCCAATACCGACTGGCCCTCAT 142706  
QY 864 CAAGCGCTGATTCGGGCGCGGATTCGCTGAGCGCGGACCAATTAACGAAGACCGCAAGA 923  
DB 142707 CAAGCGCTCATCCAAAGC-----CTGCTGTATCGCGCGGAGGAAGGAAGCGCGAGA 142760

QY 924 AGCCATGAAATGCTGGCCGAAACATACGTGGTGTGACGTGAAAGTCTGGCGGC 983  
Db 142761 GATCCGCGGCTTCTTTGCCGACGAGTACTGGGGCCGATCCGCTACGTCCAGTT 142820  
QY 984 TAGCATGAAAGGCACTTTGAAATACGAAAAAGAGATAAACGGCGCTACCGACTTCAA 1043  
Db 142821 GGGCGAGGGGGAACCGAACCTGTGCCCCCTCGGCAAGAACCGCTCCAGTACGCCACCA 142880  
QY 1044 CACCTTCTTTCCGACGGCGCCAGCTATCCGTCTACAGCAGTGACAGTCTGTATCTGAC 1103  
Db 142881 CCTGTTTCCGCGTCCGGGGTGAACCGCCCGACCGACCGACCACTGTGATCTCAC 142940  
QY 1104 CCAGCTGAGGCGCTGGGG 1121  
Db 142941 CCAGATGGCTCGCTGGGG 142958

Search completed: September 16, 2005, 06:53:40  
Job time : 5986 secs

**This Page Blank (uspto,**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2005, 02:44:35 ; Search time 751 Seconds  
(without alignments)  
10972.410 Million cell updates/sec

Title: US-10-689-200-1

Sequence: 1 atgaataacatcatcattagatc.....ccggcgcaagtcgtgat 1392

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*  
1: geneseq1808:\*  
2: geneseq1908:\*  
3: geneseq2008:\*  
4: geneseq2018:\*  
5: geneseq2018:\*  
6: geneseq2028:\*  
7: geneseq2028:\*  
8: geneseq2038:\*  
9: geneseq2038:\*  
10: geneseq2038:\*  
11: geneseq2048:\*  
12: geneseq2048:\*  
13: geneseq2048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1392	100.0	1392	12	AD006542 Methylo
2	143	10.3	551	12	AD006544 Methylo
3	85.4	6.1	747	11	ABD14925 Pseudom
4	85.4	6.1	750	11	ABD15326 Pseudom
5	85.4	6.1	831	11	ABD15218 Pseudom
6	82.4	5.9	1371	11	ACH97397 Klebsiell
7	71.8	5.2	1209	8	ACA45921 Prokaryot
8	69	5.0	594	11	ABD15217 Pseudom
9	69	5.0	3720	11	ABD15023 Pseudom
10	62.2	4.5	1173	8	ACA24128 Prokaryot
11	59.2	4.3	576	8	ACA23622 Prokaryot
12	58	4.2	1200	8	ACA23116 Prokaryot
13	58	4.2	1200	13	AD514594 Pseudom
14	58	4.2	1250	11	ABD10532 Pseudom
15	57	4.1	1116	8	ACA4309 Prokaryot
16	50	3.6	1011	9	ADA2159 DNA encod
17	49.4	3.5	1512	13	ADT43829 Bacterial
18	48.4	3.5	773	8	ACA20473 Prokaryot
19	48	3.4	744	13	ADT44805 Bacterial
20	47.8	3.4	2000	8	ADA71938 Rice gene

21	47.8	3.4	7656	8	ACA37618 Prokaryot
22	47.8	3.4	28804	2	AAT37329 Sphingom
23	47.8	3.4	28804	2	AAT92474 Sphingom
24	47.8	3.4	28804	2	AAV99812 Sphingom
25	47.8	3.4	28804	2	AAV81474 Chromosom
26	46.2	3.3	1044	12	ADJ35031 DNA encod
27	46.2	3.3	37507	12	ADH48030 Prokaryot
28	45.6	3.3	1173	8	ACA25903 Prokaryot
29	45	3.2	8298	4	AAK72613 Human Imm
30	44.8	3.2	1323	12	ADM99142 Envlronme
31	43.4	3.1	420	11	ABD15254 Pseudom
32	43.4	3.1	531	11	ABD15367 Pseudom
33	43.4	3.1	1119	11	ABD15153 Pseudom
34	43.4	3.1	1188	8	ACA25753 Prokaryot
35	43.4	3.1	1389	11	ABD14772 Pseudom
36	43.4	3.1	5577	13	ADT44774 Bacterial
37	43	3.1	1305	8	ACA23608 Prokaryot
38	43	3.1	1308	11	ABD03949 Pseudom
39	43	3.1	1359	11	ABD03676 Pseudom
40	43	3.1	1443	11	ABD04027 Pseudom
41	43	3.1	1818	13	ADS48801 Bacterial
42	42.8	3.1	516	11	ABD10614 Pseudom
43	42.8	3.1	651	11	ACH94942 Klebsiell
44	42.6	3.1	1149	11	ABD03368 Pseudom
45	42.6	3.1	1387	11	ABD03299 Pseudom

## ALIGNMENTS

RESULT 1	AD006542	standard; DNA; 1392 BP.
ID	AD006542	
XX	AD006542	
AC	29-UTR-2004	(first entry)
XX		
DE	Methylomonas sp 16a nitrogen transporter nrtA gene.	
XX		
KW	ds; gene; C1 metabolizing bacteri; gene expression regulation;	
XX	nitrogen transporter; nrtA.	
XX		
OS	Methylomonas sp.; 16a.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1392
FT		/*tag= a
FT		/product= "nitrogen transporter nrtA"
FT		/partial
FT		/note= "no stop codon"
XX		
PN	MO2004037998-A2.	
XX		
PD	06-MAY-2004.	
XX		
PF	21-OCT-2003; 2003WO-US033698.	
XX		
PR	21-OCT-2002; 2002US-0419872P.	
XX		
PA	(DUFO ) DU PONT DE NEMOURS & CO E I.	
XX		
PI	Dicosimo DU, Picataggio SK, Seip JE, Ye RW, Wang T, Ni H,	
XX		
DR	WPI; 2004-375911/35.	
XX	P-FSDB; AD006543.	
XX		
PT	Promoter regions responsive to presence of nitrate, acidic pH, elevated	
PT	temperatures or highly expressed in presence of methane or methanol	
PT	useful for driving expression (especially for producing zeaxanthin) in C1	
XX	metabolizing bacteria.	
XX		
PS	Claim 16; Page 69-70; 83pp; English.	

XX The present invention relates to a method of expressing a coding region  
 CC of interest in a CI metabolizing bacteria. Also provided are protein,  
 CC coding and promoter sequences from *Methylobacterium* sp. 16a which respond to  
 CC various metabolic and growth conditions, which can be used in the method  
 CC of the invention. The promoters are useful for driving expression of a  
 CC number of coding regions (especially for producing zeaxanthin) in CI  
 CC metabolizing bacteria under controlled conditions. The present sequence  
 CC is a *Methylobacterium* sp. coding sequence.

XX Sequence 1392 BP; 347 A; 381 C; 405 G; 259 T; 0 U; 0 Other;

Query Match 100.0%; Score 1392; DB 12; Length 1392;  
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 1392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAACCATCATTTGATGAGCTCGAAGAAATTTGTTATGACGTTATCGCTTGGCTA 60  
 DB 1 ATGAAACCATCATTTGATGAGCTCGAAGAAATTTGTTATGACGTTATCGCTTGGCTA 60  
 QY 61 GCGGTTGGGGTTGACGATGCGCCGAGATGCGGGGCAATTGGCAAGCTGGAAAGAA 120  
 DB 61 GCGGTTGGGGTTGACGATGCGCCGAGATGCGGGGCAATTGGCAAGCTGGAAAGAA 120  
 QY 121 GATTTGAAATTCGCTTCAATCAAGCTCACGCAATGAGCGCGCTGGCGGTGGCCGCGAA 180  
 DB 121 GATTTGAAATTCGCTTCAATCAAGCTCACGCAATGAGCGCGCTGGCGGTGGCCGCGAA 180  
 QY 181 AAAGGCTTTCTGAGAGACGAGGGCTGTTGTCGCACTGGAAGCGGCAACTGGAAG 240  
 DB 181 AAAGGCTTTCTGAGAGACGAGGGCTGTTGTCGCACTGGAAGCGGCAACTGGAAG 240  
 QY 241 GTGGTATGATGATGAGGTCGTGAAATGGCGAACTGGAAGCGGCTGGCACTGCTGGCGCGG 300  
 DB 241 GTGGTATGATGATGAGGTCGTGAAATGGCGAACTGGAAGCGGCTGGCACTGCTGGCGCGG 300  
 QY 301 CCGTTAGCGGCGACGCTTGGCTTGGCGACCAAGCGCGATATCGAGGTGCGTTCAAGATG 360  
 DB 301 CCGTTAGCGGCGACGCTTGGCTTGGCGACCAAGCGCGATATCGAGGTGCGTTCAAGATG 360  
 QY 361 GGGTTCAAGCGGCAAGCGGATTAACGGTGTCCATGAAATCTGGCATCAATGAAAGCCGAAC 420  
 DB 361 GGGTTCAAGCGGCAAGCGGATTAACGGTGTCCATGAAATCTGGCATCAATGAAAGCCGAAC 420  
 QY 421 ATACCGCTGGAAGCGGCTTAAACCGGTGATCCGATCAAGGAGATTAATCTTAAACCGGTC 480  
 DB 421 ATACCGCTGGAAGCGGCTTAAACCGGTGATCCGATCAAGGAGATTAATCTTAAACCGGTC 480  
 QY 481 GTGGAATAATACAAAGCCGAGGCAAGCGGTTCAATATGCGCATGACCTTCCGCGCGGA 540  
 DB 481 GTGGAATAATACAAAGCCGAGGCAAGCGGTTCAATATGCGCATGACCTTCCGCGCGGA 540  
 QY 541 TCTCAACAATCAAACTGCGTTACTGGCTGGCGGCTGGCGGTATCAATCTTGGCTACTAT 600  
 DB 541 TCTCAACAATCAAACTGCGTTACTGGCTGGCGGCTGGCGGTATCAATCTTGGCTACTAT 600  
 QY 601 TCGCGCCGCAAGACATTTCCGCGCAATCGGCGCAAGCGCTTGTGTGCGGTGACCCCG 660  
 DB 601 TCGCGCCGCAAGACATTTCCGCGCAATCGGCGCAAGCGCTTGTGTGCGGTGACCCCG 660  
 QY 661 CCGCGCAAAATGCGGTGCAAGCGTGAAGCGGCAAGCAATTTTGGTTATTTGCGTGGCGAG 720  
 DB 661 CCGCGCAAAATGCGGTGCAAGCGTGAAGCGGCAAGCAATTTTGGTTATTTGCGTGGCGAG 720  
 QY 721 CCGTGAACCAAGAGCGGCTGTTCAAGGCGATAGGCGGTGCGGTATCAACCGATGAAGAA 780  
 DB 721 CCGTGAACCAAGAGCGGCTGTTCAAGGCGATAGGCGGTGCGGTATCAACCGATGAAGAA 780  
 QY 781 CTCTGGAAGGACACGCGCGGAAAAAGTCTTGGCGGTGACCAAAACAATAGGCGGAGAAATAC 840  
 DB 781 CTCTGGAAGGACACGCGCGGAAAAAGTCTTGGCGGTGACCAAAACAATAGGCGGAGAAATAC 840  
 QY 841 CCCAACCTTATCTGGCGGTGACCAAGCGGCTGATTTGGGCGCGCATCTGGCTGACGCGC 900

DB 841 CCCAACCTTATCTGGCGGTGACCAAGCGGCTGATTTGGGCGCGCATGCTGGACGCGC 900  
 QY 901 GACATTAACAGAACCCGCAAGAGCCATTCGAATGCTGGCGCAAAAACAATATCGTGGT 960  
 DB 901 GACATTAACAGAACCCGCAAGAGCCATTCGAATGCTGGCGCAAAAACAATATCGTGGT 960  
 QY 961 GCTGACGTGAAGAGTCTGGCGGTGACATGAAAGCGCATTTTGAATACGAAAGACGAT 1020  
 DB 961 GCTGACGTGAAGAGTCTGGCGGTGACATGAAAGCGCATTTTGAATACGAAAGACGAT 1020  
 QY 1021 AAAGCGGCGCTACCGGACTTCAACACTTCTTTGCGCACGCGCGCAAGCTATCGTCTAC 1080  
 DB 1021 AAAGCGGCGCTACCGGACTTCAACACTTCTTTGCGCACGCGCGCAAGCTATCGTCTAC 1080  
 QY 1081 AGCAGTGAAGTCTGGTATCTGACCCAGCTGAGGCGCTGGGCGATATCAATTAATTCAA 1140  
 DB 1081 AGCAGTGAAGTCTGGTATCTGACCCAGCTGAGGCGCTGGGCGATATCAATTAATTCAA 1140  
 QY 1141 CCGGACAACTGGTATCTGGATATACGCGCAAGAAAGCTACCGCGCGGACATCTATCTGCGC 1200  
 DB 1141 CCGGACAACTGGTATCTGGATATACGCGCAAGAAAGCTACCGCGCGGACATCTATCTGCGC 1200  
 QY 1201 GCGGCGAAAGAACTGGTGCAGAAAGGCAAGGCGCAAGGACTTCCCTGCGCATAC 1260  
 DB 1201 GCGGCGAAAGAACTGGTGCAGAAAGGCAAGGCGCAAGGACTTCCCTGCGCATAC 1260  
 QY 1261 TCGATCAAGCGCTCGCAGAAATTTCTTCATCCAGCAAAAGTCCGTTGATGCCAAGAGCC 1320  
 DB 1261 TCGATCAAGCGCTCGCAGAAATTTCTTCATCCAGCAAAAGTCCGTTGATGCCAAGAGCC 1320  
 QY 1321 AACGATTAACCTGCGCAAGTTTGGCATAGGTCTGAAGGCAAGCAAAAGCTGACCGGCGGC 1380  
 DB 1321 AACGATTAACCTGCGCAAGTTTGGCATAGGTCTGAAGGCAAGCAAAAGCTGACCGGCGGC 1380  
 QY 1381 AAGGTCTGGAT 1392  
 DB 1381 AAGGTCTGGAT 1392

RESULT 2  
 AD006544  
 ID AD006544 standard; DNA; 551 BP.  
 XX  
 AC AD006544;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE *Methylobacterium* sp. 16a nitrogen transporter *ntrA* gene promoter.  
 XX  
 KW de; gene; CI metabolizing bacteria; gene expression regulation;  
 XX nitrogen transporter; *ntrA*; promoter.  
 OS *Methylobacterium* sp.; 16a.  
 XX  
 PN WO2004037998-A2.  
 XX  
 PD 06-MAY-2004.  
 XX  
 PF 21-OCT-2003; 2003WO-US033698.  
 XX  
 PR 21-OCT-2002; 2002US-0419872P.  
 XX  
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.  
 XX  
 PI Dicosimo DJ, Picataggio SK, Seip JB, Ye RW, Wang T, Ni H;  
 DR MPI; 2004-375911/35.  
 PT Promoter regions responsive to presence of nitrate, acidic pH, elevated  
 PT temperatures or highly expressed in presence of methane or methanol  
 PT useful for driving expression (especially for producing zeaxanthin) in CI  
 PT metabolizing bacteria.

XX Claim 37; Page 72; 83pp; English.  
PS  
XX  
CC The present invention relates to a method of expressing a coding region  
CC of interest in a C1 metabolizing bacteria. Also provided are protein,  
CC coding and promoter sequences from *Methylobacterium* sp. 16a which respond to  
CC various metabolic and growth conditions, which can be used in the method  
CC of the invention. The promoters are useful for driving expression of a  
CC number of coding regions (especially for producing zeaxanthin) in C1  
CC metabolizing bacteria under controlled conditions. The present sequence  
CC is a *Methylobacterium* sp. promoter sequence.  
XX  
SQ Sequence 551 BP; 139 A; 119 C; 135 G; 158 T; 0 U; 0 Other;  
Query Match 10.3%; Score 143; DB 12; Length 551;  
Best Local Similarity 100.0%; Pred. No. 7e-27;  
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAACCATCTTATGATGAGCTCGAGAAATTGTTATGACGTTATGCGCTTGGCTA 60  
DB 409 ATGAAACCATCTTATGATGAGCTCGAGAAATTGTTATGACGTTATGCGCTTGGCTA 468  
QY 61 GCCGTTGGGTTGACGATGCGCCGATGTCGGGCACTGGCAAGTGAAGAA 120  
DB 469 GCCGTTGGGTTGACGATGCGCCGATGTCGGGCACTGGCAAGTGAAGAA 528  
QY 121 GATTGAATTGCGCTTCATCA 143  
DB 529 GATTGAATTGCGCTTCATCA 551  
RESULT 3  
ABD14925/C  
ID ABD14925 standard; DNA; 747 BP.  
XX  
AC ABD14925;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE *Pseudomonas aeruginosa* polynucleotide #13529.  
XX  
KM Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
KM antibacterial.  
OS  
XX *Pseudomonas aeruginosa*.  
XX  
PN US651795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR MPI; 2003-615309/58.  
DR P-PDB; ABO81354.  
XX  
PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 13529; 455pp; English.  
XX  
CC The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,

CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
CC of *Pseudomonas* species using biotech technology. Sequences ABD1397-  
CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
XX  
SQ Sequence 747 BP; 114 A; 234 C; 281 G; 118 T; 0 U; 0 Other;  
Query Match 6.1%; Score 85.4; DB 11; Length 747;  
Best Local Similarity 51.3%; Pred. No. 7.5e-12;  
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;  
QY 651 GGTGACCCCGCCCGCCCAATGCGCTTCAGCTGGAAGCCGCGACATTTGCGTTATG 710  
DB 559 GGTGTTTCAACCGCCCGCAGATGTCGACATCTTCAGCCGCGCGATCGACGCGTTCTG 500  
QY 711 CGTGGCGAGCCGTGGAACAGAGCCCGTGTTCAGAGGCAATAGGCGTGGTGATGAC 770  
DB 499 CGCGCGCGCCCTTGCGGCGCCCTTGCGCTGGAACAGGCGCAAGGCTTACCATCGCCAC 440  
QY 771 CGATGAAGAACTCTGGAAGACACCGCCGGAAGAGTCTTCGCGGTGACCAATGCGC 830  
DB 439 CAGCCAGCGATCTGCGCGGACACCGGAAAGTCTTCGCGACATACCGCGCTTCGT 380  
QY 831 CGAAGAAATACCCCAACACTATCTGCGCGTGAACAGAGGCGTGAATCGGCGCGATCTG 890  
DB 379 CGACGCTTACCGCAACCGCCCGCGCTGTGATGAGC--GTCCTCGACCGCCAGCGC 323  
QY 891 GCTGAGCGCGCAATTAACAAGACCGCAAGAGGACCATCGAAATGCTGCGCGCAAAAC 950  
DB 322 CTTCATCGAACAAGAACCGGAGAACCGCTGCGACCGGCACTATACGCGCGCGCA 263  
QY 951 ATACGTGCGTGTGACGTGGAAGTCTGCGCGGTGACATGAACGCACTTTCGATACGA 1010  
DB 262 CTATGTCAGCGCTCGCTCGCGCGGATCCAGCGCGCTTCTTCGCGGTGACAGAGACG 203  
QY 1011 AAAAGACGTAAGAGCGCGCTACCGGACTTCAACACTTCTTTCGCAAGCGCGCACTA 1070  
DB 202 CCTCGGCAACGCGCTGCGAGACCCACATCCGCTGCGCTTCTTCGCGCGAGAGCTCA 143  
QY 1071 TCGGTCTCT--ACAGAGTGAAGTCTGTATCTGACCGAGCTGAGCGCGTGGCATGAT 1127  
DB 142 CGGCGCTGCGCTCTCGGATGCGATGCTTTCATGACCGCAAGTTCGCGCGCTGCT 83  
QY 1128 CAATGAA 1134  
DB 82 GCGCGAA 76  
RESULT 4  
ABD15326  
ID ABD15326 standard; DNA; 750 BP.  
XX  
AC ABD15326;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE *Pseudomonas aeruginosa* polynucleotide #13930.  
XX  
KM Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
KM antibacterial.  
OS  
XX *Pseudomonas aeruginosa*.  
XX  
PN US651795-B1.  
XX  
PD 22-APR-2003.

```
XX 18-FEB-1999; 99US-00252991.
PF 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO81755.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 13930; 455bp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
XX Sequence 750 BP; 115 A; 293 C; 237 G; 105 T; 0 U; 0 Other;
SQ
Query Match 6.1%; Score 85.4; DB 11; Length 750;
Best Local Similarity 51.3%; Pred. No. 7.5e-12;
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;
QY 651 GGTGACCCCGCCGCAATATCGCTCAAGCTGGAAGCCGACCACTTTTCGTTATTG 710
DB 54 GGTGTTTCAACCGCGCCAGATGTGCGCATCTCCAGCGCGCGATGACGCGTTCTG 113
QY 711 CGTGGCGAGCCCTGGAACGAGAGCGCGTTTCAAGGGCATAGCGCTGCGGTATCAC 770
DB 114 CGCCGGCGGCGCTCTGGGGCGCCCTGCGCCGTGGACGAGGGCCAGGGCTTACCATCGCCAC 173
QY 771 CGATGAAGAACTGTGAAGGACGCGCGGAAAAGTCTTCGGCGGTGACCAACAAATGGGC 830
DB 174 CAGCCAGCGCATTTGGCCGAGCAACCGGAAAGTCTTGGGACATACCGCGCTTCTGT 233
QY 831 GGAGAAATACCCCAACACCTATCTGCGCGGTGACCAAGCGCGTGAATTCGGGCGCGATCTG 890
DB 234 CAGACGCTTACCGGAACCGCGCGCGCGGTGATGATGCG---GTCCTGAGACGCGCGCG 290
QY 891 GCTGAGCGCGGACATTAACAAGACCGGAGGAGGACCATGGAATCTGGCGCAAAAACA 950
DB 291 CTTCATCTGAACAGAACGCGGAGAACCGCTGGGACCGGCAACTATAGCGCGCGCA 350
QY 951 ATACGTCGGTGTGAGTGAAGTGTGGGCGCTACATGAAGGACATTTGGAATACGA 1010
DB 351 CTATGTGACGCTCCGCTCGGCGGATCCAGCCGCGCTTCTTGGGCGGCTACAGAGACG 410
QY 1011 AAAAGAAGATTAACGCGCGCTACCGGACTTCAACACCTTCTTTTGGCAGCGCGCACTTA 1070
DB 411 CCTGGGCAACGCGTGGACGAGACCCACATCGCGCTTCTTAAGCGGAGGGAATCA 470
QY 1071 TCCGTCT--ACAGCATGTCAGTGTGATCTGACCCAGCTGAGCGCTGGGGCATGAT 1127
DB 471 CGGCGCTGCTCGGATGTCATGTGTTTCATGACCCAGATTCCGCGCGTGGGGCTGCT 530
```

```
QY 1128 CAATGAA 1134
DB 531 GCGCGAA 537
XX
RESULT 5
ABD15218
ID ABD15218 standard; DNA; 831 BP.
XX
AC ABD15218;
XX
XX 29-JUL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polynucleotide #13822.
XX
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX antibacterial.
XX
XX Pseudomonas aeruginosa.
XX
XX US6551795-B1.
XX
XX 22-APR-2003.
XX
XX 18-FEB-1999; 99US-00252991.
XX
XX 18-FEB-1998; 98US-0074788P.
XX 27-JUL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
DR P-PSDB; ABO81647.
XX
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
XX Disclosure; SEQ ID NO 13822; 455bp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using biochip technology. Sequences ABD01397-
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
XX Sequence 831 BP; 124 A; 327 C; 259 G; 121 T; 0 U; 0 Other;
SQ
Query Match 6.1%; Score 85.4; DB 11; Length 831;
Best Local Similarity 51.3%; Pred. No. 7.8e-12;
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;
QY 651 GGTGACCCCGCCGCAATATCGCTCAAGCTGGAAGCCGACCACTTTTCGTTATTG 710
DB 100 GGTGTTTCAACCGCGCGAGATGTGCGCATCTCCAGCGCGCGCATGACGCGTTCTG 159
QY 711 CGTGGCGAGCGCTGGAACGAGAGCGCGTGTTCAGGGCATAGGCGTGGCGGTATCAC 770
DB 160 CGCCGGCGGCGCTCTGGGGCGCGCTGGCGGTGACGAGGGCCAGGGCTTACCATGCGCAC 219
```



CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation; or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the best compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1209 BP; 229 A; 395 C; 376 G; 209 T; 0 U; 0 Other;

Query Match 5.2%; Score 71.8; DB 8; Length 1209;  
 Best Local Similarity 53.8%; Pred. No. 3e-08;  
 Matches 148; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 102 TGGCAAGCTGSAAGAAGATTGAAATTCGGCTTCACTACGACATGGCGCC 161  
 DB 45 TGAACGCCCGSAATGAACAGCTGAGCGGCTTATGAGCACTGACGACCTGCCGCC 104  
 QY 162 GCTGGCGGTGCGCCGCGAAGAGCTTCTTCGAGAGCAGAGGCGCTTTCGTGCACTGGA 221  
 DB 105 GCTGGGTGCTGCTGCCACCCAGGGGTTTGGCCAGCCTTACGGCTTAACCTGA 164  
 QY 222 AGCGACGCGCAACTGGAAGGTGTGATGATAGAGGTCTGTAATGCGAACTGACGGCTC 281  
 DB 165 AGCGACGACCTCTCTGGCGCGGCTGCTGACAGGCTGTACAGCGGCACTTACAGCGGC 224  
 QY 282 GCACATGCTGGCGCGCGCGCTTACCGCCAGCGCTTGGCTTGGGACCAAGGCCGATAT 341  
 DB 225 ACATACCTGTATGTTTATCTATCCGTTGAGAGCTGGGCAATGCGCGCGCGCGAC 284  
 QY 342 CGAGTGCCTTACAGCATGGGCTTCAACGGCAAGC 376  
 DB 285 CGACATGCGCATCTCATGGGCTCAACCAAGAACG 319

RESULT 8  
 ABD15217  
 ID ABD15217 standard; DNA; 594 BP.

XX ABD15217;  
 XX 29-JUL-2004 (first entry)  
 XX Pseudomonas aeruginosa polynucleotide #13821.

DE Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 XX antibacterial.  
 KW

OS Pseudomonas aeruginosa.

XX US651795-B1.

XX 22-APR-2003.

XX

PF 18-FEB-1999; 99US-00252991.

XX

PR 18-FEB-1998; 98US-0074788P.

PR

PR 27-JUL-1998; 98US-0094190P.

XX

(GENO-) GENOME THERAPEUTICS CORP.

PA

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX

DR MPI; 2003-615309/58.

DR

F-PSDB; AB081646.

PT

Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 useful as molecular targets for diagnostics, prophylaxis and treatment of  
 pathological conditions resulting from bacterial infection.

XX

PS Disclosure; SEQ ID NO 13821; 455bp; English.

XX

The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPRO at  
 CC segdata.uspro.gov/sequence.html

XX Sequence 594 BP; 102 A; 208 C; 196 G; 88 T; 0 U; 0 Other;

Query Match 5.0%; Score 69; DB 11; Length 594;  
 Best Local Similarity 52.2%; Pred. No. 1.3e-07;  
 Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 84 GCCCGATGTCGGGGGAGTTGGCAAGCTGSAAGAAGATTGAAATTCGGCTTCACTCA 143  
 DB 204 GGCTTGGGGCGGGGCTTCCAGACGCAACGGAAGAGCGCGCTGACATGGCTTTCATGGC 263  
 QY 144 GCTACCGACATGGGCGCGCTTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 203  
 DB 264 GCTGACGATTCGGGCTGCTGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 323  
 QY 204 CTTGTTGCTGCACTGSAAGCGCAGGCCCAACTGSAAGGTGTGATGATGATGATGATGAT 263  
 DB 324 CTTACAGCTCAATCTCAGGCGCGCAACGCTCTGGGCGCACCTTGGCGCGCAAGCTCTCG 383  
 QY 264 TGGCGAAGTGAAGCGGCTGCGACATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 323  
 DB 384 CGGCGAGCTGATGCT 443  
 QY 324 CGGCGCAAGCGCGGATATCGAGTGCCTTCAAGCATGGGCTTCAACGGCAAGC 376  
 DB 444 CGGCGCGAGCGCGCGCAAGAGATGCGGTGCTCATGGAAGCTGTCACAAAGC 496

RESULT 9  
 ABD15023/C  
 ID ABD15023 standard; DNA; 3720 BP.

XX ABD15023;  
 XX 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polynucleotide #13627.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 KW

KM antibacterial.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN US6551795-B1.  
 XX  
 PD 22-APR-2003.  
 XX  
 PF 18-FEB-1999; 99US-00252991.  
 XX  
 PR 18-FEB-1998; 98US-0074788P.  
 XX  
 PR 27-JUL-1998; 98US-0094190P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
 XX  
 DR WPI; 2003-615309/58.  
 XX  
 DR P-RSDB; AB081452.  
 XX  
 PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX  
 PS Disclosure; SEQ ID NO 13627; 455bp; English.  
 XX  
 CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biotech technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 XX  
 SO Sequence 3720 BP; 576 A; 1147 C; 1326 G; 671 T; 0 U; 0 Other;  
 Query Match 5.0%; Score 69; DB 11; Length 3720;  
 Best Local Similarity 52.2%; Pred. No. 2.4e-07;  
 Matches 153; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
 QY 84 GCCCGATGTCGGGCGAGTGGCAAGTGGGAAAGAAAGATTGAAATTCGGCTTCAATCA 143  
 DB 676 GGGCTTGGGCGGGGGGTTCCACGACCGGAAAGAGCGCCCTGGAATCGTTTCAATGCG 617  
 QY 144 GCTCAACGACATGATGCGCGCTGGCGGCGCGCGGAAAGAGCTTCTTGAAGGACGAGG 203  
 DB 616 GCTGACCGATGCGCGCTGCTGATCGTGGCGGCGGACCGAGGCTTGGCCAGCCCTACG 557  
 QY 204 CTTGTTCTGCAACTGGAAGCGGAGCGCAACTGGAAGGTTGATGATGAGTGGTCTGTA 263  
 DB 556 CCGACGCTCAATCTGACGCGGCAACGCTCTGGGCGACCTTGGGAGCAAGCTGCTCAG 497  
 QY 264 TGGCGCAATGGAAGCGGCTCGCAATGCTGGGCGCGCGCGCTTGGCGGACGCTTGGCTT 323  
 DB 496 CGGCGAGCTGGAATGCGCGGCAAGTGGCTTGTCTTACGGGCGTCAAGCTCGGCGCT 437  
 QY 324 CGGACCAAGGCGGATATGAGGTGCGGCTTCAAGTGGGCTTCAACGCGCAAG 376  
 DB 436 CGGCGGACGCGGCGGCAAGGAGATGGCGGCTGCTCAATGGAGCTGTCCAAACG 384

XX  
 AC ACA24128;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #5785.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 OS  
 OS Borrelia cepacia.  
 XX  
 PN WC0200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haseelbeck R, Ohlsen KL, Zyckind JW;  
 PI Wall D, Tremwick JD, Carr GT, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 XX  
 DR P-RSDB; AB020258.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 11998; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SO Sequence 1173 BP; 180 A; 396 C; 404 G; 193 T; 0 U; 0 Other;  
 Query Match 4.5%; Score 62.2; DB 8; Length 1173;



Best Local Similarity 58.4%; Pred. No. 9.4e-06;  
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```
QY 131 TCGGCTTCAATCAAGCTCAGCAGATGGCGCGCTGGCGGCGCCGCAAAAAGGCTTCT 190
    |||||
DB 116 TCGGCTATCTGCGCGATCAGCAGCGCGCGCTGCTGTGCGCCCAACAACGGCTATT 175
    |||||
QY 191 TCGAGGACGAGGCGCTGT---TCGTGCAACTGGAAGCGGCAACTGGAAGGTGTGA 247
    |||||
DB 176 TCGGCTCCGAAAGGCTGACGCTGAGACCAAGCTGCTGCTGAGCTGAGCGCCAGCTCG 235
    |||||
QY 248 TGGATAGGCTGTGAAATGGCGAAGTGGACCGCTCGACATGCTGGCGCGCGCTTAG 307
    |||||
DB 236 TCGAGGCGTTCTGTCCGCGCAAGTCAAGTCTGTGACCTGTGCTGCGCGATGACGCTGT 295
    |||||
QY 308 CGGCGACGCTTGGCTTGGGACCAAGCGCGATTCGAGG 346
    |||||
DB 296 GGGCGGCTACGCGACCGCGCGCGCGCAAGGTGTGG 334
    |||||
```

## RESULT 11

ACA23622  
ID ACA23622 standard; DNA; 576 BP.

XX ACA23622;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #5279.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KM drug design; gene.

XX Borrelia cepacia.

OS WO20027183-A2.

XX 03-OCT-2002.

PD 21-MAR-2002; 2002MO-US009107.

PF 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyekind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RB, Xu HH;

XX P-PSDB; ABU19752.

XX WPI; 2003-029926/02.

XX Claim 14; SEQ ID NO 11492; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide(s) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway

CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 576 BP; 84 A; 206 C; 194 G; 92 T; 0 U; 0 Other;

Query Match 4.3%; Score 59.2; DB 8; Length 576;

Best Local Similarity 57.9%; Pred. No. 4.5e-05;  
Matches 125; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

```
QY 131 TCGGCTTCAATCAAGCTCAGCAGATGGCGCGCTGGCGGCGCCGCAAAAAGGCTTCT 190
    |||||
DB 116 TCGGCTATCTGCGCGATCAGCAGCGCGCGCTGCTGTGCGCCCAACAACGGCTATT 175
    |||||
QY 191 TCGAGGACGAGGCGCTGT---TCGTGCAACTGGAAGCGGCAACTGGAAGGTGTGA 247
    |||||
DB 176 TCGGCTCCGAAAGGCTGACGCTGAGACCAAGCTGCTGCTGAGCTGAGCGCCAGCTCG 235
    |||||
QY 248 TGGATAGGCTGTGAAATGGCGAAGTGGACCGCTCGACATGCTGGCGCGCGCTTAG 307
    |||||
DB 236 TCGAGGCGTTCTGTCCGCGCAAGTCAAGTCTGTGACCTGTGCTGCGCGATGACGCTGT 295
    |||||
QY 308 CGGCGACGCTTGGCTTGGGACCAAGCGCGATTCGAGG 343
    |||||
DB 296 GGGCGGCTACGCGACCGCGCGCGCGCAAGGTGTGG 331
    |||||
```

## RESULT 12

ACA42316  
ID ACA42316 standard; DNA; 1200 BP.

XX ACA42316;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #23973.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KM drug design; gene.

OS Pseudomonas aeruginosa.

XX WO20027183-A2.

XX 03-OCT-2002.

PD 21-MAR-2002; 2002MO-US009107.

PF 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyekind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
 XX WPI: 2003-029926/02.  
 DR P-PSDB; AB038446.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 30186; 1766bp; English.  
 XX  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 1200 BP; 190 A; 420 C; 409 G; 181 T; 0 U; 0 Other;  
 SQ  
 Query Match 4.2%; Score 58; DB 8; Length 1200;  
 Best Local Similarity 56.2%; Pred. No. 0.00012;  
 Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

XX  
 AC ADS14594;  
 XX  
 DT 16-DEC-2004 (first entry)  
 XX  
 DE Pseudomonas aeruginosa quorum sensing controlled gene PA2328, SEQ ID 149.  
 XX  
 KW Cytostatic; Anti-HIV; Antimicrobial; Immunosuppressive;  
 KW quorum sensing signalling; bacterium; quorum sensing controlled gene;  
 KW biofilm-associated disease; cystic fibrosis; AIDS; middle ear infection;  
 KW acne; periodontal disease; gene; ds; PA2328.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN WO2004083385-A2.  
 XX  
 PD 30-SEP-2004.  
 XX  
 PF 11-MAR-2004; 2004WO-US007467.  
 XX  
 PR 14-MAR-2003; 2003US-00389647.  
 XX  
 PA (IOWA ) UNIV IOWA RES FOUND.  
 XX  
 PI Greenberg EP, Schuster M, Lostrich C;  
 XX  
 DR WPI: 2004-709932/69.  
 XX  
 PT Identifying a modulator of quorum sensing signalling in bacteria, useful  
 PT for treating a biofilm-associated disorder, comprises contacting the cell  
 PT with a quorum sensing signal molecule in the presence and absence of a  
 PT test compound.  
 XX  
 XX Disclosure; SEQ ID NO 149; 233bp; English.  
 XX  
 XX The present invention relates to a method for identifying a modulator of  
 CC quorum sensing signalling in bacteria. The method comprises: providing a  
 CC cell that comprises a quorum sensing controlled gene (ADS1446-ADS14798),  
 CC where the cell is responsive to a quorum sensing signal molecule such  
 CC that a detectable signal is generated; contacting the cell with a quorum  
 CC sensing signal molecule in the presence and absence of a test compound;  
 CC and detecting a change in the detectable signal. The method and modulator  
 CC identified by the method are useful for treating a biofilm-associated  
 CC disease or disorder, e.g. cystic fibrosis, AIDS, middle ear infections,  
 CC acne, periodontal disease, catheter-associated infections, and medical  
 CC device-associated infections. Note: The sequence data for this patent was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 1200 BP; 190 A; 420 C; 409 G; 181 T; 0 U; 0 Other;  
 SQ  
 Query Match 4.2%; Score 58; DB 13; Length 1200;  
 Best Local Similarity 56.2%; Pred. No. 0.00012;  
 Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

Db 371 GCGGCTCGGGGCTCACGGTGGCTCCGAGATC 402

RESULT 14  
ID ABD10532 standard; DNA; 1260 BP.  
XX  
XX ABD10532;  
AC  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX  
XX Pseudomonas aeruginosa polynucleotide #9136.  
DE  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KM  
XX  
XX antibacterial.  
OS  
XX  
XX Pseudomonas aeruginosa.  
PN  
XX  
XX US6551795-B1.  
PD  
XX  
XX 22-APR-2003.  
PF  
XX  
XX 18-FEB-1999; 99US-00252991.  
PR  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR  
XX  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
PI  
XX  
XX WPI: 2003-615309/58.  
DR  
XX  
XX P-PSDB; ABO76961.  
PT  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT  
XX  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT  
XX  
XX pathological conditions resulting from bacterial infection.  
PS  
XX  
XX Disclosure; SEQ ID NO 9136; 455pp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biotech technology. Sequences ABD01397-  
CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
XX  
XX Sequence 1260 BP; 202 A; 450 C; 423 G; 185 T; 0 U; 0 Other;

Query Match 4.2%; Score 58; DB 11; Length 1260;  
Best Local Similarity 56.2%; Pred. No. 0.00012;  
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

QY 131 TCGGCTTCATCAAGCTCAACCGATAGCGCGCGTGGCGCGCCGCAAAAAGCTTCT 190  
DB 197 TCGGCTTACCTGCGAATCAACCGATAGCGCGCGCTTGTGTGGCCATGCCAACGCGCTGT 256  
QY 191 TCGAGGACGAGCGGCTGTTCGTGCACTGGAAGCGAGGCC--AACTGGAAGTGTGTA 247  
DB 257 TCGAGGCGGAGGCGATTCAGGCGGAAACGCGCGTGTCTTGTGCTACTGGGCGCAGGTAA 316  
QY 248 TGGATAGGCTGTGAAATGGCGAACTGACCGGCTCGCAATGTCTGAGCGCGCGCGCTTAG 307

Db 317 TCGAGGCGGTTCATCTCCGCGCAGGTCAACGTTCACCTGCTGCGGATACCGTCT 376

QY 308 CGGCGACGCTTGAGCTTCCGGACCAAGCGCGATATGAGTGCCTTCAGCATGGGCTTCA 367  
DB 377 GGGCC-----CGTACGCGACAGCAAGGTGCGCGCGCGCATGTGCTTGGAACCACTCG 430

QY 368 ACGGCAAGCGATTCAGGCTGTGTCATGAAATC 399  
DB 431 GCGGCTCGGGGCTCACGGTGGCTCCGAGATC 462

RESULT 15  
ID ACA44309 standard; DNA; 1116 BP.  
XX  
XX ACA44309;  
AC  
XX  
XX 19-JUN-2003 (first entry)  
DT  
XX  
XX Prokaryotic essential gene #25966.  
DE  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KM  
XX  
XX drug design; gene.  
KW  
XX  
XX Pseudomonas putida.  
OS  
XX  
XX WO200277183-A2.  
PN  
XX  
XX 03-OCT-2002.  
PD  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR  
XX  
XX 06-SEP-2001; 2001US-00948993.  
PR  
XX  
XX 25-OCT-2001; 2001US-0342923P.  
PR  
XX  
XX 08-FEB-2002; 2002US-00072851.  
PR  
XX  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;  
PI  
XX  
XX Wall L, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
PT  
XX  
XX WPI: 2003-029926/02.  
DR  
XX  
XX P-PSDB; ABU40439.

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 32179; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 623 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-regulated gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence data for this target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1116 BP; 215 A; 361 C; 342 G; 198 T; 0 U; 0 Other;

Query Match 4.1%; Score 57; DB 8; Length 1116;

Best Local Similarity 54.1%; Pred. No. 0.00021;

Matches 139; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

```
QY 131 TCGGCTTCATCAAGCTCAGCATGAGCGCGCTGGGCGTGGCCGCAAAAGCTTCT 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 56 TCGGCTACCTGCCATCAAGCTCAGCGCGCTGGTGGCAATTAACAGGCTGT 115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 TCGAGACGAGGGCT--GTTCTGCACTGGAAGCGCAGGCCCACTGGAAGTGTGA 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116 TCGAGGCCGGAAGCATCAAGCGCGCGCTGGTGGCAAGCTGGGCGAGGTGA 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 TGGATAGGGTCTGTAATGGCGAAGCTGCAAGCTGCTGGCGCGGCGCTTAG 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 176 TCGAGGCATTCATCTCCGCGCAGGTCAATGTCATCACTGCTTCCGCGATGACGTAT 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 CGGCGACGCTTGGCTTGGGCAACCAAGCCGATATCGAGGTGGCGTTAGCATGGGCTTCA 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 236 GGGCGCGCTTATGGCAGCAAGGTGCCAGCCCAAGGTGGCGCTGGAACCAAGTGGCGGTT 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 ACGGCAACGCGATTACG 384
    ||||| ||||| |||||
Db 296 CGGGCTTGACGTGGCG 312
    ||||| ||||| |||||
```

Search completed: September 16, 2005, 05:13:49  
Job time : 759 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2005, 04:12:19 ; Search time 4286 Seconds

(without alignments)  
12362.451 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392  
Sequence: 1 atgaataacatcatgtagtc.....ccggcgcaagtcgtgcat 1392

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.4	6.7	1872	9	CL982367 OaIFSC047
2	87.6	6.3	575	8	A2935061 Bf_Ba000
3	82.4	5.9	800	8	B2575941 msh2_470.
4	61.8	4.4	1085	7	CF161028 B0692F03-
5	58.6	4.2	1085	8	B2548675 pac81-60
6	49.2	3.5	1035	8	B2563077 pac81-164
7	48.6	3.5	1407	8	B2572733 pac81-2775
8	48.4	3.5	788	8	B2550725 pac81-60
9	47.8	3.4	459	9	AL056652 Drosophila
10	47.2	3.4	691	6	CA718866 wkm2n.pko
11	47.2	3.4	691	6	CD894397 G118_126C
12	47.2	3.4	824	7	CN129290 RHO1_34
13	46	3.3	623	6	CA699640 wlk8_pko0
14	45.6	3.3	469	6	CA066446 SCEQAD101
15	45.6	3.3	500	6	CA069366 SCSFAD106
16	45.6	3.3	602	6	CA069388 SCSFAD106
17	45.6	3.3	616	6	CA219519 SCSFAD106
18	45.6	3.3	645	6	CA212608 SCSFAD111
19	45.6	3.3	683	6	CA264998 SCSFAD111
20	45.6	3.3	708	6	CA069230 SCSFAD105
21	45.6	3.3	758	6	CA217278 SCSFAD112
22	45.6	3.3	840	6	CA174317 SCSFAD101
23	45.6	3.3	914	6	CA106423 SCSFAD101
24	45.4	3.3	564	2	BE361762 Dcl_82_E0

25	45.4	3.3	660	6	CD231149 SSI_8_F06
26	45.4	3.3	663	6	CB925634 ABA1_22_G
27	45.4	3.3	811	7	CN141741 WOUNDI_1
28	45	3.2	503	8	CC351449 OGIAE24TV
29	45	3.2	824	8	B2628435 1h60d07_9
30	45	3.2	852	8	CC0313120 PUGH08TB
31	44.8	3.2	568	2	BE775653 MY-05-C-0
32	44.8	3.2	590	2	BE776036 MY-10-B-0
33	44.6	3.2	472	4	Bj210804 B210804
34	44.6	3.2	580	4	Bj218083 B218083
35	44.6	3.2	604	6	CA253508 SCSFAD110
36	44.6	3.2	686	6	CA134878 SCSFAD101
37	44.6	3.2	695	9	CL164460 SC18P101
38	44.6	3.2	942	8	B2575093 msh2_430.
39	44.4	3.2	930	9	CL969992 OaIFSC019
40	44.2	3.2	547	8	BH633034 1007058E1
41	44.2	3.2	672	6	CA194351 SCRFBS102
42	44.2	3.2	788	9	CC700945 OCVBN08TH
43	44.2	3.2	1483	8	B2573056 msh2_2923
44	44.2	3.2	1605	8	B2573144 msh2_2963
45	44	3.2	424	9	CL794328 OR_CBA000

## ALIGNMENTS

RESULT 1	CL982367	1872 bp	DNA	linear	GSS 21-SEP-2004
LOCUS	OaIFSC047480	Oryza sativa	Express Library	Oryza sativa	(indica
DEFINITION	Cultivar-group) genomic, genomic survey sequence.				
ACCESSION	CL982367				
VERSION	CL982367.1	GI:52419216			
KEYWORDS	GSS.				
SOURCE	Oryza sativa (indica cultivar-group)				
ORGANISM	Oryza sativa (indica cultivar-group)				
REFERENCE	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.				
AUTHORS	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis				
TITLE	Unpublished (2004)				
JOURNAL	Contract: Chen Chen				
COMMENT	Department of Bioinformatics Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.				
FEATURES	Location/Qualifiers				
source	1..1872				
	/organism="Oryza sativa (indica cultivar-group)"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:39946"				
	/clone_lib="Oryza sativa Expressed Library"				
	/note="Oryza sativa exon trapped genomic sequences"				
ORIGIN					
Query Match	6.7% Score 93.4; DB 9; Length 1872;				
Best Local Similarity	51.9%; Pred No. 5.1e-14;				
Matches 261; Conservative 0; Mismatches 236; Indels 6; Gaps 2;					
QY	633 CCGAGCGCTTGTGTGCGTACCCCGCGCGGCAATGCCGTCCAGCGTGAAGCCGG				692
DB	552 CCATGACGCGGACGAGCGTGTGTCGCCCGCGGAGATGTGATGAGACATCGCATTTG				611
QY	693 CACCAATTTTCGTTATTTGCGTGGCGAGCGCGTGAACAGCAGCGCGTTCAGAGGCAT				752

```

Db      612 CAACATGCTGGGCTTTTGGCTGCTGAGCGCTGGAACGGAGGGCGATTACGACCGCAT 671
Qy      753 AGCGCTGCGCGTATATACCGATGAAGAATCTTGAAAGAGACCGCGGAAAAAGCTTTCCG 812
Db      672 TGGCTTCAACCGCTGCCACGTCGACATCTTGCGGAGATCATCTGAAAAAATTTCTCCG 731
Qy      813 CGGACCAAAACAAATGGCGGAGAAATACCCCAACACTTATCTGGCGGATGCCAAGCGCT 872
Db      732 TACGCTCGGACCTGGGTGGAGAAAATCCGACACCGCCGCGCTGTGAGCGCT 791
Qy      873 GATTCGGGCGCGATCTGGCTGAGCCGACAAATAAACAAGAACCGAACCCATCGA 932
Db      792 GATGGAAGCCGCGCTGATGAGGCGTCCCGGAAAAAACAACG---TGAACCGCGCA 848
Qy      933 AATGCTGGCGCAAAAACAAATACGTGGTGTGACGTGGAAGTCTTGCGCGCTATGCAATGA 992
Db      849 GATCCTCTCGCGCGCGGCGCTCAACATGCAAGAACAGTATCTACCGCGCGAATGCT 908
Qy      993 CGGCACTTTGAAATACGAAAAAGACGATAAAGCGGCGCTACCGGACTTCAACACTTTCT 1052
Db      909 CGGGGAATACGACAAAGCGCTGGGGGACGCGCTGGAGAGCGCACCCGATCTTTCTT 968
Qy      1053 TCGCCACG---CGCCAGCTATCCGTCTACAGACAGTGCAGTCTGTATCTGACCCAGCT 1109
Db      969 CAACGAGGGTGGCGCTGATATCCGTACCTCTCCGACGGCATGTGTTCTTAACCAAGTT 1028
Qy      1110 GAGCGCCTGGGCGCATGATCAATG 1132
Db      1029 CCGCGCGCTGGGCTTCTTAAG 1051

```

## RESULT 2

A2935061 575 bp DNA linear GSS 24-APR-2001  
 LOCUS B2935061  
 DEFINITION B\_Ha0003E04r B. japonicum BAC library Bradyrhizobium japonicum genomic, genomic survey sequence.

ACCESSION A2935061  
 VERSION A2935061  
 KEYWORDS GSS  
 SOURCE Bradyrhizobium japonicum  
 ORGANISM Bradyrhizobium japonicum  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1 (bases 1 to 575)  
 Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,  
 Goicechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.,  
 A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome

AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,  
 Goicechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.,  
 TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum  
 genome

JOURNAL Genome Res. 11 (8), 1434-1440 (2001)  
 MEDLINE 21376150  
 PUBMED 11483585

COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: twing@clemson.edu  
 Class: BAC ends  
 High quality sequence stop: 535.

FEATURES  
 source  
 1..575  
 Location/Qualifiers  
 /organism="Bradyrhizobium japonicum"  
 /mol\_type="genomic DNA"  
 /strain="USD110"  
 /db\_xref="taxon:375"  
 /lab\_host="E. coli"  
 /clone\_lib="B. japonicum BAC library"  
 /note="Vector: pindig0536; Site\_1: HindIII"

## ORIGIN

Query Match 6.3%; Score 87.6; DB 8; Length 575;

```

Best Local Similarity 61.1%; Pred. No.1.5e-12;
Matches 160; Conservative 0; Mismatches 99; Indels 3; Gaps 1;
Qy      131 TGGCTTATCAAGCTCACCGACATGGCGCGCTGGCGCGCCGCGGAAAAAGCTTCT 190
Db      214 TGGGTTCAATTCCTGCTGATCGATGCGCGCGCTATGTCGCTCTGCAACAAGGCTTTA 273
Qy      191 TCGAGACGAGAGGCGCTTTCGTGCAACTGGAAGCGCAGGCGCAACTGGAAGGTGATAG 250
Db      274 CCGCGCGGAAAGGCTTCGACATGCTGAACTGTGCGGAGAGTCTCTGTGTCACAGTCCGG 333
Qy      251 ATAGGATCGTGAATGGCAATGACAGCGCTGCGACATGCTGGCGCGCGCGCTTACGCG 310
Db      334 ACAAGCTCAATATCGCGCTGTTGACGCGCGCCCATCTGCTGGCGCGCGGCGATCGGCT 393
Qy      311 CCAAGCTTGGCTTCCG---CACCAAGCGCATATGAGAGTCCGTTCAAGCATGGCTTCA 367
Db      394 CTTCACTCGGCGCTCGGCACATGCAAGGTGCGATGCGCGCGCTTCAATCTCGGCATCA 453
Qy      368 ACGGACGCGGATTAAGGATGTC 389
Db      454 ACGGACGCGGATTAAGGATGTC 475

```

## RESULT 3

BZ575941 800 bp DNA linear GSS 17-DEC-2002  
 LOCUS BZ575941/C  
 DEFINITION msh2\_470.x2 msh Pseudomonas aeruginosa genomic clone msh2\_470,  
 genomic survey sequence.

ACCESSION BZ575941  
 VERSION BZ575941.1  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 800)  
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haetings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press

TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library

JOURNAL Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216934  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Haetings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library

JOURNAL Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216934  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1..800  
 Location/Qualifiers  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="MSH"  
 /db\_xref="taxon:287"  
 /clone\_lib="msh2\_470"  
 /note="Environmental isolate. Whole genomic shotgun  
 library."

ORIGIN  
 Query Match 5.9%; Score 82.4; DB 8; Length 800;  
 Best Local Similarity 50.7%; Pred. No.4.2e-11;  
 Matches 225; Conservative 0; Mismatches 216; Indels 3; Gaps 1;

```

Qy      651 GGTGACCCCGCGCGCAATGCGCTCCACGCTGGAAGCGGACCATTTTCGTTATG 710
Db      526 GGTGTTTCAACCGCGGAGATGTGCGGCATCTCCAGGCCGCGCATGACGGGTTCTG 467
Qy      711 CGTCGGCAGCGCGTGAACAACAGAGCGCGTGTTCAGGCGATAGGCGGTGATGATAC 770
Db      466 CGCCGCGGAGACCTTGCGCGCGCTGCGCTGAGACAGGCGCATTCACCATCGCAC 407

```



QY 771 CGATGAAGACTCTGAGAGCAACCGCGAAGAAAGTTTGGGCTGACCAAGATGGC 830  
 Db 406 CAGCCAGCGATCTGGCGGACCAACCGGAAAGTCTTGGCACTACCGCGCTTCTGT 347  
 QY 831 GGAGAAATACCCCAACCACTATCTGCGGTGACCAAGCGCGTGAATTCGGCGCGATCTG 890  
 Db 346 CGAGCGCTACCCGAAACACCGCCCGCGCTGTCTATGCG---GCTCTGACACCGCAGCCG 290  
 QY 891 GCTGAGCGCGACATTAACAAGAACCGCAAGAAAGCCATGAAATGCTGGCGCAAAAACA 950  
 Db 289 CTTTCATGCAACAGAACCGCGAGAACCGCTGGGCAACCGCGCAACTGATGACGCGCGCA 230  
 QY 951 ATACGTGCGTCTGCTGACCTGGAAGTCTGGCGGTCTAGCATGAAGGCACTTTCGAAATGCA 1010  
 Db 229 CTATGTGAGCGCTCCGCTGCGCGATCCAGCGCGCTTCTTCCGCGCTACCAAGACGG 170  
 QY 1011 AAAGAAGATAAAGCGCGCTACCGGACTTCAACACCTTCTTCCGCGCGCGCAAGCTTA 1070  
 Db 169 CCTCGGCAAGCGCTGCGAGAACCGCGATCCGCTGCGCTTCTATGCGACGCGCAAGTCAA 110  
 QY 1071 TCCGCTCTACAGACAGTGCAGTCTG 1094  
 Db 109 CCGGCGCTGCTCTCGAGGCGCTG 86

RESULT 4  
 CFI61028 406 bp mRNA linear EST 25-JUL-2003  
 LOCUS B0692F03-5 NIA Mouse Blastocyst cDNA library (Long) Mus musculus  
 DEFINITION cDNA clone NIA.B0692F03 IMAGE:30457118 5', mRNA sequence.  
 ACCESSION CFI61028  
 VERSION CFI61028.1 GI:33270577  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 406)  
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.  
 Construction of long-transcript enriched cDNA libraries from  
 submicrogram amounts of total RNAs by a universal PCR amplification  
 method  
 Genome Res. 11 (9), 1553-1558 (2001)

JOURNAL MEDLINE  
 PUBMED 21429098  
 11544199

COMMENT  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@lgsun.grc.nia.nih.gov  
 Plate: B0692 row: F column: 03  
 Seq primer: M13 Reverse  
 High quality sequence stop: 406  
 POLY-A-No.

FEATURES  
 source  
 1..406  
 location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="taxon:10090"  
 /clone="NIA:B0692F03 IMAGE:30457118"  
 /issue\_type="Blastocyst"  
 /dev\_stage="3.5-dpc"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Blastocyst cDNA library (Long)"  
 /note="Vector: pSPORT1 (Invitrogen) Site 1: SalI; Site 2:  
 NotI; Mouse cDNA project by the Laboratory of Genetics,  
 National Institute on Aging (NIA), Intramural Research  
 Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is  
 a long-transcript enriched cDNA library (Ref. Genome Res.  
 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were

extracted from a pool of 20 Blastocysts. Double-stranded  
 cDNAs were synthesized with an oligo (dT) primer  
 [Invitrogen]:  
 5'-pAGCTAGTCTAGATCGGACGCGCGCCCTTTTCTTTT-3' from  
 0.2 ug of total RNA, treated with T4 DNA polymerase, and  
 purified by ethanol-precipitation. The cDNAs were ligated  
 to lone-linker L1-SalI, purified by phenol/chloroform, and  
 separated from free linkers by centrifugation 100. Then, the  
 cDNAs were amplified by long-range high fidelity PCR using  
 Ex Taq polymerase (Takara) with a primer SalI-4-S. The  
 products were purified by phenol/chloroform and centrifuged  
 100. The cDNAs were digested with SalI and NotI enzymes  
 and cloned into SalI/NotI site of pSPORT1 plasmid vector.  
 The DH10B E. coli host was transformed with the ligation  
 mixture by the standard chemical method. The average  
 insert size is about 2.2 kb. The library was constructed  
 by Yulan Piao (NIA)."

ORIGIN  
 Query Match 4.4%; Score 61.8; DB 7; Length 406;  
 Best Local Similarity 52.5%; Pred. No. 1.5e-05;  
 Matches 160; Conservative 0; Mismatches 142; Indels 3; Gaps 1;

QY 630 CCGCGGAGAGCGCTTGTGCGGTGACCCGCGCGCGCAAAATCGCTCCAGCTGAGAGC 689  
 Db 79 CGACACCGATGTCACATCCGCCACATGCGCGCGCGCAAAATGTTCCAGCTGCGCGC 138  
 QY 690 CCGCACCATTTTGTGTTATGCTGCGCGAGCGGTGAAACAGACAGCGCGTTCAGAGG 749  
 Db 139 CGACACCATGACAGCGCTTCTCTGCGCGCGCGCGAGGACCGAGCGCGGTCTATACCG 198  
 QY 750 CATAGCGGTGCGCGGTATCATCCGATGAAGACTCTGGAAGAGACCGCGGAAAAAGTCTT 809  
 Db 199 CGTGGCGCTTATCATCATCTGACCAAGATATCTGGGAAGGCGCATCGTGCCTCGCTT 258  
 QY 810 CCGCGTGAACCAACAAATGCGCGGAGAAATACCCCAACACTTATCTGCGGTGACCAAGC 869  
 Db 259 CCGCGCTCTCAAGAAATGTTGTACACGCGTCCGAAACACTTATGCGCATTTGTGAATC 318  
 QY 870 GCTGAT---TCGGCGCGCATGCTGCTGAGCGCGCAATTAACAGAACCGCAAGAGAGC 926  
 Db 319 GATCATGATGAGAGCGCGGTGCGCGCAAGCGCGAACCAGAACAGATTGCGGAGAGC 378  
 QY 927 CATCG 931  
 Db 379 AATCG 383

RESULT 5  
 BZ548675/c 1085 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacel1-60\_1334.s1 pacel1-60 Pseudomonas aeruginosa genomic clone  
 DEFINITION pacel1-60\_1334, genomic survey sequence.  
 ACCESSION BZ548675  
 VERSION BZ548675.1 GI:27152256  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 1085)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press  
 Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu

```

FEATURES          Class: shotgun.
                   Location/Qualifiers
    source          1..1085
                   /organism="Pseudomonas aeruginosa"
                   /mol_type="genomic DNA"
                   /strain="1-60"
                   /db_xref="taxon:287"
                   /clone_1fb="pac1-60.1334"
                   /note="clinical isolate 1-60 whole genomic shotgun
                   library."

ORIGIN
Query Match          4.2%; Score 58.6; DB 8; Length 1085;
Best Local Similarity 52.2%; Pred. No. 0.00014;
Matches 153; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 84 GCCCGATGTCGGGCGAGTTGGCAAGCTGMAAAGAAATTGAAATTCGGCTTCATCA 143
    |||||
DB 463 GGCCTGGGCGGCGGTTCCGACCGCAACGGAAAGAGCGCCCTGGACATCGGTTTCATGCG 404

QY 144 GCTCACCGACATGCGCCGCTGCGGTCGGCCCGGAAAAAGGCTTCTTGAGAGACGAGG 203
    |||||
DB 403 GCTGACCGATTGCGCTCGCTGATCGTGGCCGCC-ACCGATGCTTGCAGCCCTACCG 345

QY 204 CTTGTCGTCACTGGAAGCGCAGGCAACTGGAAGTGTGATGATGAGGTGCTGAA 263
    |||||
DB 344 CTCACGCTCAATCTCAGGCGGCCAACCTGCTGGGCCACCTTGCGGACAAACTGCTCAG 285

QY 264 TGGCGAAGCTGAGCGGCTCGCACATGCTGGCGCGCGCGGTAGCGGCCAGCGTTG 323
    |||||
DB 284 CGCGAGCTGATGTCGCGCGAGTGCTGTACGGGGTGTCTACGCGGTCAAGCTCGGCT 225

QY 324 CGGCACCAAGCGCGATATCGAGTGCCTTCAGCATGCGCTTCAACGGCAAGC 376
    |||||
DB 224 CGCGGCGACGCGCAAGCGAGTGGCGGTGCTCATGCGAGTGTGCCAGAACG 172

RESULT 6
BZ563077          1035 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION      pac62-164_4075.x1 pac62-164 Pseudomonas aeruginosa genomic clone
ACCESSION      BZ563077
VERSION        BZ563077.1 GI:27186268
KEYWORDS
SOURCE
ORGANISM      Pseudomonas aeruginosa
               Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
               Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 1035)
AUTHORS      Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
               Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE        Whole-Genome-Sequence variation among multiple isolates of
               Pseudomonas aeruginosa library
JOURNAL      J. Bacteriol. (2002) In press
COMMENT      Contact: Chris K. Raymond
               Genome Center
               University of Washington
               Box 352145, Seattle, WA 98105-2145, USA
               Tel: 2062216954
               Fax: 2066857244
               Email: craymond@u.washington.edu
               Class: shotgun.
               Location/Qualifiers
    source          1..1035
                   /organism="Pseudomonas aeruginosa"
                   /mol_type="genomic DNA"
                   /strain="2-164"
                   /db_xref="taxon:287"
                   /clone_1fb="pac2-164_4075"
                   /clone_1fb="pac2-164"
                   /note="clinical isolate 2-164 whole genomic shotgun"

```

```

ORIGIN          library."
Query Match          3.5%; Score 49.2; DB 8; Length 1035;
Best Local Similarity 55.9%; Pred. No. 0.05;
Matches 113; Conservative 0; Mismatches 88; Indels 1; Gaps 1;

QY 658 CGCGCGCGCAAAATGCGCGTCCACGCTGGAAGCGCGGACCATTTGGTATTATGCGTGGC 717
    |||||
DB 448 CCACCGGCGCACATGTCGGCCATCTCCAGCGCGGCGCATGACGAGTTCTGCGCGGCG 507

QY 718 -GAGCGTGAACACGACAGCGCGTGTCAAGGCGCTAGCGCGTGCCTGATCAACGATGA 776
    |||||
DB 508 GCACCTCTGGGCGCGCTCGCGCGTGAACAGGCGCAGGCGCTTCAACATGCCACAGGCA 567

QY 777 AGAATCTGGAAGGACAGCGCGGAAAAAGTCTTTCGCGCTGACCAACAATGCGCGAGAA 836
    |||||
DB 568 GGCAGTTTGGCCCGGACACCGGAAAAAGGCGCTCGAAACTACCGCGCTTGTGACGC 627

QY 837 ATACCCCAACACTATCTGCGG 858
    |||||
DB 628 CTACCCGAACACCGCGCGCG 649

RESULT 7
BZ572739          1407 bp      DNA      linear      GSS 17-DEC-2002
LOCUS          mh2_2775.y2 mh Pseudomonas aeruginosa genomic clone mh2_2775,
DEFINITION      genomic survey sequence.
ACCESSION      BZ572739
VERSION        BZ572739.1 GI:27207800
KEYWORDS
SOURCE
ORGANISM      Pseudomonas aeruginosa
               Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
               Pseudomonadaceae; Pseudomonas.
REFERENCE      1 (bases 1 to 1407)
AUTHORS      Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
               Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE        Whole-Genome-Sequence variation among multiple isolates of
               Pseudomonas aeruginosa library
JOURNAL      J. Bacteriol. (2002) In press
COMMENT      Contact: Chris K. Raymond
               Genome Center
               University of Washington
               Box 352145, Seattle, WA 98105-2145, USA
               Tel: 2062216954
               Fax: 2066857244
               Email: craymond@u.washington.edu
               Class: shotgun.
               Location/Qualifiers
    source          1..1407
                   /organism="Pseudomonas aeruginosa"
                   /mol_type="genomic DNA"
                   /strain="MSH"
                   /db_xref="taxon:287"
                   /clone_1fb="mh2_2775"
                   /clone_1fb="mh"
                   /note="Environmental isolate. Whole genomic shotgun
                   library."

ORIGIN
Query Match          3.5%; Score 48.6; DB 8; Length 1407;
Best Local Similarity 50.7%; Pred. No. 0.078; Indels 3; Gaps 2;
Matches 155; Conservative 0; Mismatches 148; Indels 3; Gaps 2;

QY 84 GCCCGATGTCGGGCGAGTTGGCAAGCTGMAAAGAAAGATTGAATTGCGCTTCATCA 143
    |||||
DB 330 GGCCTGGGCGGCGGTTCCGACCGCAACGGAAAGAGCGCCCTGGAATCGGTTTCATGCG 389

QY 144 GCTCACCGACATGCGCGCGCTG-GGGTGGCGCGCGGAAAAAGGCTTCTTGAGAGACGAG 202
    |||||
DB 390 GCTGACCGATTGCGCTCGCTGATGTTGCGCGCCACCGAGGCTTTCGCCCGCTACG 449

```

QY 203 GCCTGTTCTGCACTGGAAGCCGCACTGGAAGTGTGATGATAGGTGCTGTA 262  
 DB 450 GCTTCAAGCTCAATCTCAGGCGCCCAACCGTCTGGGCGACCTTGCGCGACAACTGCTCTCA 509  
 QY 263 ATGGCGAAGTGAAGGCTGCGCATGCTGCGCGCGCGCTTAGCGGCGAGCG--TTGG 320  
 DB 510 GCGGCGACAGNAGATCGNNNNNCAGTGTCTACGGGCTTGTTACCGGCGCTCCAGCTCG 569  
 QY 321 CTTGGCACCAAGCGCGATATTCAGGTGCGCTTCAGCATGCGCTTCAACGCGCAACCGCAT 380  
 DB 570 CCTCGGCGGACAGCGCGCGACGAAGGCGCGTGTCTCATGACATGCTGCGAGAACGCGAG 629  
 QY 381 TACGCT 386  
 DB 630 GCGCAT 635

RESULT 8  
 B2550725/c 788 bp DNA linear GSS 17-DEC-2002  
 LOCUS pacel-60\_2795.y2 pacel-60 Pseudomonas aeruginosa genomic clone  
 DEFINITION pacel-60\_2795, genomic survey sequence.  
 ACCESSION B2550725  
 VERSION B2550725  
 KEYWORDS B2550725.1 GI:27154306  
 SOURCE GSS.  
 ORGANISM Pseudomonas aeruginosa  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 1 (bases 1 to 788)  
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.  
 Whole-Genome-Sequence Variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press  
 Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Classes: shotgun.  
 FEATURES  
 source Location/Qualifiers  
 1..788  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="1-60"  
 /db\_xref="taxon:287"  
 /clone="pacel-60\_2795"  
 /clone\_1lb="pacel-60"  
 /note="clinical isolate 1-60 Whole genomic shotgun  
 library."  
 ORIGIN  
 Query Match 3.5%; Score 48.4; DB 8; Length 788;  
 Best Local Similarity 51.3%; Pred. No. 0.078;  
 Matches 138; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

DB 202 TCTGGCCCCCTACGCGACAGAGTCCCGCCCAAGGTGTGCTTGAACACACGTGGCG 143  
 QY 371 GCAACGCGATTACGTTGTCATGAATC 399  
 DB 142 GCTCGGGGCTCACGTTGCTCCGAATATC 114

RESULT 9  
 CNS0052P 844 bp DNA linear GSS 03-JUN-1999  
 LOCUS CNS0052P  
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #  
 BACR1P16 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL056652  
 VERSION AL056652  
 KEYWORDS AL056652.1 GI:4932342  
 SOURCE GSS.  
 ORGANISM Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 844)  
 Genoscope.  
 DIRECT SUBMISSION  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osogawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
 FEATURES  
 source Location/Qualifiers  
 1..844  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR1P16"  
 /clone\_1lb="RPCI-98"  
 /note="end : TET3"  
 ORIGIN  
 Query Match 3.4%; Score 47.8; DB 9; Length 844;  
 Best Local Similarity 17.2%; Pred. No. 0.12;  
 Matches 60; Conservative 143; Mismatches 145; Indels 0; Gaps 0;

Qy 711 CGTGGCGAGCCGTGGAAACAGAGGCGCGTTCAGAGGCATATGAGGCTGGGTATTCAC 770  
 Db 599 SRVSSCSGSSSSSSSSGCGSVRSRSGSANGVATGVSGRGGCGGCRGAGAAARADDA 658  
 Qy 711 CGATAGAACTCTGTGAGAGACACGCCGGAAGAAAGTCTTGGCGGTAC 818  
 Db 659 ABRARAKARANSNADAMGAIVSSSVSSBAGRNGRRAAAVVGVSGB 706

RESULT 10	CA718866	459 bp	mRNA	linear	EST 26-NOV-2002
LOCUS	CA718866				
DEFINITION	wk2m.pk002.c3 wk2m Triticum aestivum CDNA clone wk2m.pk002.c3 5' end, mRNA sequence.				

ACCESSION	CA71886
VERSION	CA71886.1
	GI:25440659

KEYWORDS	EST.
SOURCE	Triticum aestivum (bread wheat)
ORGANISM	Triticum aestivum

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta; Spermatophyta, Magnoliophyta, Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Triticum.

REFERENCE  
AUTHORS  
1 (bases 1 to 459)  
Tingey, S. V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,

TITLE Miao, G., Caraher, N. and Hanafey, M. K.  
JOURNAL Dupont Wheat cDNA Sequence  
COMMENT Unpublished (2002)  
Contact: Scott V. Tingey

E. I. Dupont de Nemours and Company  
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com  
Seq primer: M13.

```

FEATURES
source
location/Qualifiers
1. 459
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Spring wheat"
/db_xref="taxon:4565"
/clone="wkm2n_pk002.c3"
/tissue_type="kernel"
/lab_host="DH10B"
/clone_1b="wkm2n"
/note="Vector: pbluescript SK+; Site_1: EcoRI; Site_2:
XhoI; wheat (Triticum aestivum L.) kernel malted 175 hours
at 4 C, normalized"

```

	Query Match	3.4%	Score 47.2	DB 6	Length 459
	Best Local Similarity	50.9%	Pred. No. 0.15	Mismatches 108	Indels 0
	Matches	112	Conservative 0	Mismatches 108	Indels 0
QY	156	GGCGCCGTCGCGGTGGCGCCGAAAAGGCTTCTTCGAGGACGAGGGCTGTTGTCGA	215		
Db	48	GGCAGCATATGCGACGGCGCAGCAGAGTCGGAACAACGAGCGAGGCGGCCGACCGGGGA	107		
QY	216	ACTGAAACCGCAGCGCAACTGGAAGGTGTCATGATGGTTCGTGAATGCGCAACTGGA	275		
Db	108	CTCGGGGGCGTTCCGTTTCGCGGTAAAGCGAGGAGGGCGCCGCGATCGAGAGCCGG	167		
QY	276	CGGCTCGCACATGCTGTCGCGCGCGCCGCTTAGCGGCCGCTTGGCTTTCGACCAAGGC	335		
Db	168	CGCCTCGCGCGGAGCGGCGTCGAGTCCCTCGGAGACGAGGGTCGACGACGCGCAATCTCA	227		
QY	336	CGATATCGAGGTGCGCTTCACGATGAGGCTTCAAGCGCAAC	375		
Db	228	GGAGGCCAGTCTCTGCTTCGCGAGAGGGCTCTCTCTCAAC	267		

RESULT 11  
CD894397

LOCUS	CD894397	691 bp	mRNA	linear	EST 14-JUN-2003
DEFINITION	G118.126622R010823 G118 Triticum aestivum cDNA clone G118126C22, mRNA sequence.				
ACCESSION	CD894397				
VERSION	CD894397.1	GI:32665884			
KEYWORDS	EST.				
SOURCE	Triticum aestivum (bread wheat)				
ORGANISM	Triticum aestivum				

REFERENCE

Spematophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
1 (bases 1 to 691)

AUTHORS	Genopiante.
TITLE	Genopiante, a major partnership french program in plant genomics
JOURNAL	Unpublished (2003)
COMMENT	Contact: Genopiante

Genopiantle  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00  
Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme "Genopiantle" (<http://www.genopiantle.com>  
and <http://genopiantle-info.infobioisen.fr>).

```

FEATURES
source
location/Qualifiers
1..691
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="rectal"
/db_xref="taxon:4565"
/clone="G18126C2"
/tissue_type="grain (118 degrees per day after
pollination)"
/clone_id="G118"

```

Query Match	3.4%	Score 47.2	DB 6	Length 691
Best Local Similarity	50.9%	Pred. No. 0.16		
Matches 112; Conservative	0	Mismatches 108	Indels 0	Gaps 0

[illegible]

LOCUS	DEFINITION
CN129290	824 bp mRNA linear EST 01-APR-2004
CN129290	RHOH1_34_C09.g1_A002 Acid and alkaline-treated roots Sorghum
CN129290	bicolor cDNA RHOH1_34_C09.A002 5', mRNA sequence.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
CN129290	CN129290.1	GI:45956911	Sorghum bicolor (sorghum) EST.	Sorghum bicolor (sorghum)
			Sorghum bicolor	Ekaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 824)	Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Lim, S., Eastman, A. and Pratt, L.H.	An EST database from Sorghum: acid- and alkaline-treated roots

JOURNAL  
COMMENT

Unpublished (2003)  
Other ESTs: RH01.34 C09.b3 A002  
Contact: Cordouner-Fratt M  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpatr@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in  
the Human Genome Center, University of Tokyo Institute of Medical  
Science; plant material and RNA prepared at Texas A & M University;  
sequencing done in the Laboratory for Genomics and Bioinformatics,  
University of Georgia. Sequence ends have been trimmed to exclude  
vector and regions below phred quality 16. Three-prime sequences  
are presented as their reverse complement and have been trimmed to  
exclude polyA.

Seq primer: Sug5 (CTTGTGCTTAAGCTCG).

FEATURES  
source

Location/Qualifiers

1..824

/organism="Sorghum bicolor"

/mol\_type="mRNA"

/cultivar="BTx623"

/db\_xref="taxon:4558"

/clone="RH01.34 C09.A002"

/lab\_host="DH10B-T1 phage-resistant E. coli"

/note="Organ: Root; Vector: pME18S-FL3; Site: 1: XhoI;  
Site 2: XhoI; The library was prepared from polyA+ RNA  
from 8-day-old roots harvested from BTx623 sorghum  
seedlings grown in hydroponic culture. HCl was added to a  
pH of 3.0 to some seedlings, KOH to a pH of 9.0 for  
others. Roots were harvested 3, 12 and 27 hr after  
addition and pooled for RNA preparation. Double-stranded  
cDNA was cloned unidirectionally into different DraIII  
sites of the pME18S-FL3 vector (5-prime DraIII site is  
CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises  
the cDNA insert."

## ORIGIN

Query Match 3.4%; Score 47; DB 7; Length 824;

Best Local Similarity 52.3%; Pred. No. 0.19;

Matches 104; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 159 GCCGCTGGCGGTCGCGCGGCAAAAAGGCTTCTTGAGAGCAGAGGCGCTGTGTCGCACT 218

DB 319 GCTGCTGGCGGCTCTACTTCCACGACTGCTTCCGCAAGGCTGCGAGCGCGTGTGCTGT 378

QY 219 GGAAGCGCAGGCGCACTGGAAGTGTGATGATAGGCTGTGAATGGCAATGGACG 278

DB 379 GGAAGCGCAGGCGCAGAGAGGCGCGCGCAACCAAGTCGCTGGGCGGCTGACGT 438

QY 279 CTCGCACATGCTGGCGCGCGCGCTTAAAGCGCGCAGCGTGTGGGACCAAGGCCG 338

DB 439 CGTCGACGCGCGCCCAAGGCGCGCGCTGAGAGCGCGCTGCGCGGACCGTCTCTGCGCG 498

QY 339 TATCGAGGTGCGCTTACG 357

DB 499 CGTCGTGGCGCTGGCCACC 517

## RESULT 13

LOCUS

CA699640 623 bp mRNA linear EST 26-NOV-2002

DEFINITION wlk8.pk0022.g9 wlk8 Triticum aestivum cDNA clone wlk8.pk0022.g9 5'

ACCESSION CA699640

VERSION CA699640.1 GI:25421433

KEYWORDS EST.

SOURCE Triticum aestivum (bread wheat)

ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Poideae; Triticaceae; Triticum.  
1 (bases 1 to 623)  
Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hailey,C., Yuan,Z.,  
Mao,G., Caraher,N. and Hanafey,M.K.  
Dupont Wheat cDNA Sequence  
Unpublished (2002)  
Contact: Scott V. Tingey

Crop Genetics  
E. I. Dupont de Nemours and Company  
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com  
Seq primer: M13.

FEATURES  
source

Location/Qualifiers

1..623

/organism="Triticum aestivum"

/mol\_type="mRNA"

/cultivar="Stephens"

/db\_xref="taxon:4565"

/clone="wlk8.pk0022.g9"

/issue\_type="leaf"

/clone\_lib="wlk8"

/note="Vector: pGEM-7; Site 1: EcoRI; Site 2: XhoI; Wheat  
(Triticum aestivum L.) seedlings 8 hr after treatment with  
6-iodo-3-propyl-2-propyloxy-4(3H)-quinazolinone"

## ORIGIN

Query Match 3.3%; Score 46; DB 6; Length 623;

Best Local Similarity 46.3%; Pred. No. 0.33;

Matches 116; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 158 CGCGCTGGCGGTCGCGCGCGCAAAAAGGCTTCTTGAGAGCAGAGGCGCTGTGTCGCAAC 217

DB 51 CGAGGTGGCGGTCGCGCGCGCAGAGGCGGTGACGAGATGCGCGCTGCGCGGAGAGG 110

QY 218 TGAAGCGCAGGCGCACTGGAAGTGTGATGATGAGTGTGTAATGGCGAACTGACG 277

DB 111 CGAGCGCGATGACCCCGCGGTCGCGCAGCAGAGAGGCGGTGNGATGATGTCATGCG 170

QY 278 GCTCGCACATGCTGGCGCGCGCGCTTAAAGCGCGCAGCGTGTGGCTTGGCACAAGCGCG 337

DB 171 GCGAGAGCGCGGACCGATGACCCCGCGCGCGGAGGCGTCCAGGCGCTTACGAGCC 230

QY 338 ATATCGAGGTGCGCTTACGATGAGGCTTCAACGCGCAACCGATTAACGCTGTCGAATGA 397

DB 231 GACGTGTATGAGTGTGTGCTCTCGTGTCAAGCTAATATGAAAGCGCCGACGCGTG 290

QY 398 TCTGCGATGATGAAGCGCAATATCCGCTGGAAGCGGTAAACCGGTGATC 451

DB 291 TACCGTCAAGAGAAAGAAACATCTGCGCGCGGTTCGCGCGTTCGCGTTC 344

## RESULT 14

LOCUS

CA066446 469 bp mRNA linear EST 23-SEP-2003

DEFINITION SCEQAD1015H11.g AD1 Saccharum officinarum cDNA clone SCEQAD1015H11

ACCESSION CA066446

VERSION CA066446.1 GI:34917970

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum  
complex.

REFERENCE 1 (bases 1 to 469)

Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P  
Centro de Biologia Molecular e Engenharia Genetica



C	1	85.4	6.1	747	4	US-09-252-991A-13529	Sequence 13529, A
	2	85.4	6.1	750	4	US-09-252-991A-13930	Sequence 13930, A
	3	85.4	6.1	831	4	US-09-252-991A-13822	Sequence 13822, A
	4	82.4	5.9	1371	4	US-09-489-039A-3192	Sequence 3192, A
	5	69	5.0	594	4	US-09-252-991A-13821	Sequence 13821, A
	6	58	5.0	3720	4	US-09-252-991A-13627	Sequence 13627, A
	7	58	4.2	1260	4	US-09-252-991A-9136	Sequence 9136, A
	8	50	3.6	1011	4	US-09-328-352-3446	Sequence 3446, A
	9	48.2	3.5	1368	4	US-09-302-540-7805	Sequence 7805, A
C	10	48.2	3.5	6655	4	US-09-302-540-777	Sequence 777, A
	11	48	3.4	750	4	US-09-302-540-5934	Sequence 5934, A
	12	48	3.4	72704	4	US-09-302-540-1273	Sequence 1273, A
	13	47.8	3.4	28804	2	US-08-592-874-1	Sequence 1, A
	14	47.8	3.4	28804	3	US-09-096-942-2	Sequence 2, A
	15	47.8	3.4	28804	3	US-09-096-867-2	Sequence 2, A
	16	44.4	3.2	7218	1	US-08-232-463-14	Sequence 14, A
	17	43.8	3.1	3384	4	US-09-302-540-3760	Sequence 3760, A
	18	43.8	3.1	23333	4	US-09-302-540-1184	Sequence 1184, A
	19	43.4	3.1	420	4	US-09-252-991A-13858	Sequence 13858, A
	20	43.4	3.1	531	4	US-09-252-991A-13971	Sequence 13971, A
	21	43.4	3.1	1119	4	US-09-252-991A-13757	Sequence 13757, A
	22	43.4	3.1	1389	4	US-09-252-991A-13376	Sequence 13376, A
C	23	43.4	3.1	5523	4	US-09-302-540-2081	Sequence 2081, A
	24	43.4	3.1	5577	4	US-09-302-540-5383	Sequence 5383, A
	25	43.4	3.1	34552	4	US-09-302-540-1262	Sequence 1262, A
	26	43	3.1	1308	4	US-09-252-991A-2253	Sequence 2253, A
	27	43	3.1	1359	4	US-09-252-991A-2280	Sequence 2280, A

C	28	43	1	1443	4	US-09-252-991A-2631	Sequence 2631, App
	29	42.8	3.1	516	4	US-09-252-991A-9218	Sequence 9218, App
	30	42.6	3.1	651	4	US-09-489-039A-737	Sequence 737, App
	31	42.8	3.1	1149	4	US-09-252-991A-1972	Sequence 1972, App
C	32	42.6	3.1	1587	4	US-09-252-991A-1903	Sequence 1903, App
	33	42.6	3.1	1836	4	US-09-252-991A-2056	Sequence 2056, App
	34	42.4	3.0	430	4	US-09-489-039A-2830	Sequence 2830, App
C	35	42.4	3.0	465	4	US-09-452-991A-1459	Sequence 1459, App
	36	42.4	3.0	939	4	US-09-489-039A-2942	Sequence 2942, App
C	37	42.4	3.0	1512	4	US-09-252-991A-1551	Sequence 1551, App
	38	42.4	3.0	1677	4	US-09-252-991A-1616	Sequence 1616, App
C	39	41.8	3.0	2025	4	US-09-252-991A-10293	Sequence 10293, App
	40	41.8	3.0	2054	4	US-09-252-991A-10564	Sequence 10564, App
	41	41.6	3.0	1425	4	US-09-240-236-972	Sequence 972, App
C	42	41.6	3.0	10528	4	US-09-402-540-945	Sequence 945, App
	43	41.4	3.0	1154	4	US-09-902-540-5610	Sequence 5610, App
	44	41.4	3.0	1206	4	US-09-302-540-4100	Sequence 4100, App
C	45	41.4	3.0	11812	4	US-09-502-540-1041	Sequence 1041, App

## ALIGNMENTS

## RESULT 1

```

US-09-252-991A13529/C
: Sequence 13529, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ. ID NOS: 33142
: SEQ. ID NO 13529
: LENGTH: 747
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-13529

```

Query Match 6.1%; Score 85.4; DB 4; Length 747;

Best Local Similarity 51.3%; Pred. No. 1.3e-12;  
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2

Qy	651	GGGAGACCCCGCGCGCGCAAAATGCGTCCACGCTGGAAGCGCGGACCATTTTGGATTATG	710
Db	559	GGTGGTTTCAACCGGCGCGAGATGGTTCGACATCTCCAGGCGGGCGCATGCACGGTTCTG	500
Qy	711	CGTCGCGAGCGCGTGGAAACAGACGCGCGTTGTCAGGGCATAGCGCTGCCTGATTCAC	770
Db	499	CGCGCGGCGCCCTGGGCGCGCTTGGCGTGGACCAAGGCGCAAGGGCTTCACATCGCCAC	440
Qy	771	CGATGAAGACTCTTGGAAAGACACGCGCGGAAAAAGTTTGGCGGTGACCAACAAATGGCG	830
Db	439	CAGCCAGGGGATCTTGGCGGACCAACCGGAAAAAGGTCCTTGGGCACTAACCGCGCCTTCGT	380
Qy	831	GGAGAAATATCCCGCAACACTTATCTGGCGGATGACCAAGCGCTGATTCGGGCGCGATCTG	890
Db	379	CGAGCGCTTACCGGAACACCGCGCGCGCTGTGTATGG---GGTCTCGACGCCGACCGG	323
Qy	891	GCTGGACGCGCGACATTAACAAGAACCGCAAGAAAGCCATGAAATGCTGGCGCAAAAACA	950
Db	322	CTTTCATCGAACGAAGCGCGAGAACCGGCTGGGACCGCGCAACTGTATCAGCGGCGCGA	263
Qy	951	ATATCGTCCGTGTACACGTGGAAGTGTCTGGCGGCTTACGATAGACGCACTTTGGAATACGA	1011
Db	262	CTATGTGCAAGCTTCGCGTCCGCGGCGATTCACGCGCGCTTTCTTTCGCGCGCTTACAGGACGG	203



QY 1011 AAAAGAGATTAACGCGCGCTACCGGACTTCAACACTTTCTTTGCGCAAGCGCCAGCTA 1070  
DB 202 CCGGCGCAACGCGCTGCGAGAGACCCACATCGCGTTCGTTTACGCCGAGCGGAATCA 143  
QY 1071 TCCGTCCT--ACAGAGTGCAGTCTGTGATCTGACCCAGCAGTGAAGCGCTGGGGCATGAT 1127  
DB 142 CCGGCGCTGCGCTCGGAGTGCATGTGTTATGACCAGATTCCGCGCTGGGGCGCTGCT 83  
QY 1128 CAATGAA 1134  
DB 82 GCGCGAA 76

## RESULT 2

US-09-252-991A-13930  
; Sequence 13930, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13930  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13930

Query Match 6.1%; Score 85.4; DB 4; Length 750;  
Best Local Similarity 51.3%; Pred. No. 1.3e-12;  
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;

QY 651 GGTGACCCCGCGCCCAATGCGCTGCAGCTGGAAGCGGACACATTTTGGTTATG 710  
DB 54 GGTGTTCCACCGCGCGAGATGCTGCGCATCTTCAGCGCGCGCATGACGCGTTCTG 113  
QY 711 CGTGGCGAGCCGCTGGAACAGCAGGCGCGTTCAGAGGCAATAGGCGTCCGTCATCAG 770  
DB 114 CCGCGCGCGCGCTTGGGGCGCGCTGCGTGCAGCAGGGCGAGGCTTACCATGCGCCAG 173  
QY 771 CGATGAAGAACTCTGGAAGACAGCGCGAAAAAGTCTTCGCGGTGACCAACATGGGC 830  
DB 174 CAGCCAGCGATCTGCGCCGAGCACCGGAAAAAGTCTCGGACATACCGCGCTTCGT 233  
QY 831 GGAAGAAATACCCCAACCTATCTGCGCGTGAACCAAGCGCGTGAATTCGCGCGCATCTG 890  
DB 234 CAGCGCTACCGCAACCGCGCGCGCTGCTGATGAGC--GTCCTTGACGCGCAGCG 290  
QY 891 GCTGAGCGCGCAATTAACAAGAACCGCAAGAGCATGAAATGCTGGCGCAAAAAACA 950  
DB 291 CTTCATGCAACAGAACCGCGAGAACCGCTGCGCAACCGCAACTATCAGCGCGCGCA 350  
QY 951 ATACGTGCTGCTGACGTGAAGAGTGTGCGCGCTAGCATGAACGGCACTTTGGAATACGA 1010  
DB 351 CTATGTCGACGCTCCGCTGCGCGCATTCACGCGCGCTTCTTCGCGCGCTACAGAGCG 410  
QY 1011 AAAAGAGATTAACGCGCGCTACCGGACTTCAACACTTTTTCGCAAGCGCCAGCTA 1070  
DB 411 CCGGCGCAACGCGCTGCGAGAGCCACATCGCTGCGTTTCTACGCGAGCGGAAAGTCAA 470  
QY 1071 TCCGTCCT--ACAGAGTGCAGTCTGTATCTGACCCAGCTGAAGCGCTGGGGCATGAT 1127  
DB 471 CCGGCGCTGCGCTCGGATGCGATGTGTTCATGACCCAGTTCCGCGCTGGGGCGCTGCT 530  
QY 1128 CAATGAA 1134  
|||

DB 531 GCGCGAA 537

## RESULT 3

US-09-252-991A-13822  
; Sequence 13822, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13822  
; LENGTH: 831  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13822

Query Match 6.1%; Score 85.4; DB 4; Length 831;  
Best Local Similarity 51.3%; Pred. No. 1.3e-12;  
Matches 250; Conservative 0; Mismatches 231; Indels 6; Gaps 2;

QY 651 GGTGACCCCGCGCCCAATGCGCTGCAGCTGGAAGCGGACACATTTTGGTTATG 710  
DB 100 GGTGTTCCACCGCGCGAGATGCTGCGCATCTTCAGCGCGCGCATGACGCGGTTCTG 159  
QY 711 CGTGGCGAGCCGCTGGAACAGCAGGCGCGTTCAGAGGCAATAGGCGTCCGTCATCAG 770  
DB 160 CCGCGCGCGCGCTTGGGGCGCGCTGCGTGCAGCAGGGCGCAAGCTTCAACATGCGCAC 219  
QY 771 CGATGAAGAACTCTGGAAGACAGCGCGAAAAAGTCTTCGCGGTGACCAACATGGGC 830  
DB 220 CAGCCAGCGATCTGCGCCGAGCACCGGAAAAAGTCTTCGACATACCGCGCTTCGT 279  
QY 831 GGAAGAAATACCCCAACCTATCTGCGCGTGAACCAAGCGCGTGAATTCGCGCGCATCTG 890  
DB 280 CAGCGCTACCGCAACCGCGCGCGCTGCTGATGAGC--GTCCTTGACGCGCAGCG 336  
QY 891 GCTGAGCGCGCAATTAACAAGAACCGCAAGAGCATGAAATGCTGGCGCAAAAAACA 950  
DB 337 CTTCATGCAACAGAACCGCGAGAACCGCTGCGCACCGCGCAACTGATACGCGCGCGA 396  
QY 951 ATACGTGCTGCTGACGTGAAGAGTGTGCGCGCTAGCATGAACGGCACTTTGGAATACGA 1010  
DB 397 CTATGTCAGCGCTCCGCTGCGCGCATTCAGCGCGCTTCTTCGCGCTACAGAGCG 456  
QY 1011 AAAAGAGATTAACGCGCGCTACCGGACTTCAACACTTTTTCGCAAGCGCGCAGCTA 1070  
DB 457 CCGGCGCAACGCGCTGCGAGAACCCACATCGCGCTGCTTCTACGCGCGAGAGTCAA 516  
QY 1071 TCCGTCCT--ACAGAGTGCAGTCTGTATCTGACCCAGCTGAAGCGCTGGGGCATGAT 1127  
DB 517 CCGGCGCTGCGCTCGGATGCGATGTGTTATGACCCAGTTCCGCGCTGGGGCGCTGCT 576  
QY 1128 CAATGAA 1134  
DB 577 GCGCGAA 583  
|||

## RESULT 4

US-09-489-039A-3192  
; Sequence 3192, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

```

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709, 2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117, 747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 3192
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-3192

```

Query Match	Similarity	5.9%	Score 82.4	DB 4	Length 1371
Best Local	Similarity 56.2%		Pred. No. 9.7e-12		
Matches 155	Conservative 0		Mismatches 121		Indels 0; Gaps 0
QY	652	GTGACCCCGCGCCGCAATGCGCTGTCCACGCTGTGAAGCCGCGACCATTTTCGGTATTATTC	711		
Db	694	GTGTGTCCGCGCGCCGCGCAATGTGTATGAAACATGCCCATCTGCGAATATGACGGGCTTTTTCG	753		
QY	712	GTCGGCGAGCCGTGTGAACACGAGCCCGTGTTCAAAGGCGATAGGCGGTGATTCACC	771		
Db	754	GTCGGCGAGCGGTGTGAACGCGCCGCGCTATTAAACGACGATTCGCTTCAACCGCGGCACT	813		
QY	772	GATGAAGAATCTGTGAAGACACGCGCGGAAAAAGTCTTCGTGCGGTGATACCAAAATAGGCGC	831		
Db	814	TCCCAAGGATATCTGTGCGCCGAGCATCTCGGAAAAAGTGTGGGCAACCGGTGCGACTGGGTG	873		
QY	832	GAGAAATACCCCAACACCTATCTGGCGGTTGACCAAGGCGGCTGATTCGGGCGCGCATCTTGG	891		
Db	874	GAAAGCAACCCGAACACCGCCGCGCGCTGTGTGGCGGCTTGATGAAAGCGCAGCGCTGG	933		
QY	892	CTGACGCGCGCAATATACAGAAGCGCAAGGAAGCC	927		
Db	934	ATCGCCGCTTCCGCGGAGACACCTCGGAGACGCGCC	969		

```

RESULT 5
US-09-252-991A-13821
; Sequence 13821, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS.
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13821
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-13821

```

	Query Match	5.0%	Score 69	DB 4	Length 594
	Best Local Similarity	52.2%	Pred. No. 2.4e-08		
	Matches 153	Conservative 0	Mismatches 140	Indels 0	Gaps 0
Qy	84	GCCCGATGTCGGGCGAGTTGGCAAGCTTGAAAGGAAATTTGAATTCGGCTTCATCA	143		
Db	204	GGCCTGGGCGCGCGGTTCCGAGCGCAGAAAGAGCCCTGCAATCGGTTTCAATGGC	263		
Qy	144	GCTCACCGACATGGCGCGCGCTGGCGGTCGCGCCGCAAAAAGCTTCTTTCAGACAGAGG	203		
Db	264	GCTACCGAATTCGGCGCTCGCTGATGTCGTGGCGGCACCAAGGAGCTTTCGCCACGCCCTAAGG	323		
Qy	204	CTGTGTCGCACTCGAAGCGCAGGCCAATCGAAGGTGTGATGATTAAGGTCGTGAA	263		

Db 324 CCTCACGCTCAATCTTCAGGCGCCCAACCGTCTCTGGGCGACACTTTCGGGACAAGTGTCTCAG 3833

Qy 264 TGGCAACTGAGCGGCTCGCACATGTGGCGCGCGCGCCGTTAGCGGCGACGCTTGGCTT 3223

Db 384 CGGCAGGTGGATGTCGGCGGACATGCTGTACCGGGCTTGTCTACGGGTGCACGCTCGGCCCT 4433

Qy 324 CGGCACCAAGGCGCATATGAGATGCGCTTACGACATGGGCTTCAACGGGCAAG 376

Db 444 CGGCGGCGACGCGGCGCAAGCGAATGCGGTGCTCTCATGGACATGTGCCAAACG 496

RESULT 6  
 US-09-252-991A-13627/C  
 ; Sequence 13627, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 13627  
 ; LENGTH: 3720  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-13627

	Query Match	Similarity	5.0%;	Score 69;	DB 4;	Length 3720;	
	Best Local	Similarity	52.2%;	Pred.	4.7e-08;		
	Matches	153;	Conservative	0;	Mismatches	140;	Indels 0; Gaps 0;
OY	84	GCCCCGATGTGGGGSCAGTTGGCAAGCTGGAAAAAGAAAGATTGGAATTTGGGCTTCAATCAA	143				
Db	676	GGCTTGGGCGCGCGGTTCACACGACCGGAAAAAGAGCCCTGGACATCGGTTTCATGAGC	617				
OY	144	GCTCACGACATGGCGCGCTGGCGGTGGCGCGCGCGCAAAAAGCTTCTTCAGAGACGAGG	203				
Db	616	GCTGACCGGATTCGGGCTTCGTGTGATCGTGGCGCGCACCCAGGGGCTTGGCCAGCCCTACGG	557				
OY	204	CTGTTCGTGCACTGGAAAGCGCAGGCCAACTGGAAAGTGTGATGATAGGTGCTGAA	263				
Db	556	CTTCACGCTCAATCTCAGGGCGCCAAACGTCCTGGGCCACCTTGCGCACAAGCTGCTCAG	497				
OY	264	TGGCGAATCGAAGCGGCTCGCACATGCTGGCGCGCGCGCCCTTATAGCGGCCAGCGTTGGCTT	323				
Db	496	CGGGAGACTGGATGCGCGGACATGCGCTGTAAAGGGCTTGTACGGCGTCCAGCTCGGCGCT	437				
OY	324	CGGACACCAAGCGCGATTCGAGGTGCCGTTCAAGATGAGGCTTCAAGCGCAAG	376				
Db	436	CGGGGAGAGCGCGGCAAGCGAGATGGCGGTCTCATGGAAGCTGTGCCAAAAAG	384				

```

RESULT 7
US-09-252-991A-9136
: Sequence 9136, Application US/09252991A
: Patent No. 6551795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190

```

; PRIOR FILING DATE: 1998-07-27  
 ; SEQUENCE OF SEQ ID NOS: 33142  
 ; SEQ ID NO 9136  
 ; LENGTH: 1260  
 ; TYPE: DNA  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-9136

Query Match 4.2%; Score 58; DB 4; Length 1260;  
 Best Local Similarity 56.2%; Pred. No. 2,4e-05;  
 Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

QY 131 TCGGCTTCACTCAAGCTCACCAGCATGCGCCCGCTGCGCGCCGAAAAAGGCTTCT 190  
 DB 197 TCGGCTTCACTGCGCATCACCAGCGCCCGCTTGTGTGCGCCCAAGCGCTTCT 256  
 QY 191 TCGAGAGCAGGAGCGCTTGTGTGCACTGAAAGCGAGCC--AACTGAAAGTGTGTA 247  
 DB 257 TCGAGGCGCAGGAGCATCCAGGCGCAACGCGGTGCTGTGCGTACTGGGCGCAGGTGA 316  
 QY 248 TGGATAGGCTCGCAATGCGCACTGAGCGCTCGCACTGCGCGCGCGCGCGCTTAG 307  
 DB 317 TCGAGGCGCTTCACTTCCGCGCAGGTCAAGCTCATCCACTGCTGCTGCGCATGACCTCT 376  
 QY 308 CGGCGCAGCTTGGCTTCCGCGCAGGCGCGATTCGAGGTGCGCTTCAAGCATGCGCTTCA 367  
 DB 377 GGGCC-----CGCTACGCGCAGCAAGTCCCGCCCAAGGTGCTGCGCTTGAACAGCTCG 430  
 QY 368 ACGGCAACGCGATTAACGCTGTCCATGAATC 399  
 DB 431 GCGGCTCGGGGCTACGCTGTGCTCCGAGATC 462

RESULT 8  
 US-09-328-352-3446  
 ; Sequence 3446, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; FILE REFERENCE: BAUMANNIT FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; PRIOR FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 3446  
 ; LENGTH: 1011  
 ; TYPE: DNA  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-3446

Query Match 3.6%; Score 50; DB 4; Length 1011;  
 Best Local Similarity 51.8%; Pred. No. 0.0028;  
 Matches 113; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

QY 658 CCGCGCGCGCAATGCGCTGCAAGCGCGGACCATTTTGGTTATTTGCGTGGC 717  
 DB 496 CCGCGCGCGTATATGTCGAAGCTTTAGATTAACATGATGATGATTTTGTGTGGGA 555  
 QY 718 GAGCGGTGAACACGAGGCGCTGTTCAAGGCGATAGGCGCGGTGATCACCGATGAA 777  
 DB 556 GAGCCTTGGAAATACATCAAGTGAATTTTATGTTTAAATGATGCTGCTCAAGTCAA 615  
 QY 778 GAACCTGGAAGACACGCGCGGAAAAAGCTTTCGCGGTACCAACAAATGCGCGAGAAA 837  
 DB 616 GACATTATTTCCAAATGTGGCAATAAAGTTCTCGAGTACTCAAGAAATGGGCTGAACA 675  
 QY 838 TACCCCAACACTATCTGCGGTGACCAAGCGCTGAT 875  
 DB 676 CATCTCACACCTTAGTTCCTTACACGCAATTTAT 713

RESULT 9

US-09-902-540-7805  
 ; Sequence 7805, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 7805  
 ; LENGTH: 1368  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-7805

Query Match 3.5%; Score 48.2; DB 4; Length 1368;  
 Best Local Similarity 44.3%; Pred. No. 0.0095;  
 Matches 197; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 195 GGACGAGGCGCTTGTGCACTGGAAGCGGAGCCCACTGGAAGGTGTGATGATAG 254  
 DB 93 GGTCAAGTGCAGATTCATGAACCCGCGCGCTTCCATCAAGAGCCGATGACCTCTCAAT 152  
 QY 255 GGTCTGATATGCGCAATCTGACGCGCTCGCACTGCGCGCGCGCGCTTATGCGCGCAG 314  
 DB 153 CCTGAGAAAGCGCGAGCGCGGAGGGAAGCTCAAGCCCGCGGCAACATGCTGAGAAAC 212  
 QY 315 GCTTGGCTTCCGACCAAGCGCGATATGAGGTGCTGCTTCAAGATGGCTTAAAGCGCAA 374  
 DB 213 GTCCGCGCAACCGGCAATGAGGCTGCGCTGCGCGCGCGCTTCAAGGCTCAAGATGCAAT 272  
 QY 375 CGCGATTCGAGTGTCAATGAATCTGATCATGATGAAGCGGACCAATACCGCTGAAAGG 434  
 DB 273 CTTCACCAATGCCGCAAGATGTCTTGAAGATCAACCGCTTCAAGGCGCTGAGGCC 332  
 QY 435 CGGTAACCGGTGATCCGATCAAGGAGATTTATTAACCGGTGCTGAAAAATACAA 494  
 DB 333 GCAAGTGTGATGAGCGCGGAGAGAGTGCAGGCGGAGGAGCCGCGGAGCTACTAGAGAC 392  
 QY 495 AGCGGAAGCAAGCGCTTCAATATGCGATGACTTCCCGCGGATCTCACAAATCAA 554  
 DB 393 GGCAGAGGCGCTGACCGCGGAGAGCGCGGCGCTTCAATGCTCAACAGTACCAACAGCC 452  
 QY 555 ACTGCTTACGAGTGGCGGCGGCTGATCAATCTGAGCTACTATTGCGCGCGCGAGGA 614  
 DB 453 CGACAACTTCAGAGGCGCACTACACACACCGGTCTGATGATTTCAAGACAGACGAGG 512  
 QY 615 CATTTCCGCGCAATATGCGCGAGAC 639  
 DB 513 CAAGTTCACTACTTGTGCGGGG 537

RESULT 10  
 US-09-902-540-777/c  
 ; Sequence 777, Application US/09902540  
 ; Patent No. 6833447  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; PRIOR FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 777  
 ; LENGTH: 6655  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-777

Query Match 3.5%; Score 48.2; DB 4; Length 6655;  
 Best Local Similarity 44.3%; Pred. No. 0.017;  
 Matches 197; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

195 GGAAGAGGAGGCTGTTCTGTCGAACTGGAAGCGCAGGCCCACTGGAAGAGTGTGATGATAG 254  
 DB 4813 GGTCAAGTGTGAATTCATGAACCCGGCGCTTCATCAAGACCGCATGCGCTTCAAT 4754  
 QY 255 GGTGTGATGTGGAACTGGAACGCGCTTCGACATGTGGCGCGCGCTTAAAGCGGCGAG 314  
 DB 4753 CCTCGAAGAGGCGAGCGGAGGGAAGCTCAAGCCGCGGCACTGTCGAGAAAC 4694  
 QY 315 CGTTGGCTTCGGACCAAGCGCGATATCGAGTGCCTTCAAGATGGGCTTCAAGCGCA 374  
 DB 4693 GTCCGGCAACACCGGATGTGGCGCTGGCGCGCGCTCAAGGCTTCAAGTGCAT 4634  
 QY 375 CGGATTAACGGTGTCCCAATGAATCTGGCATCAGATGAAGCGCACTACCGCTGGAAG 434  
 DB 4633 CTTCACATCTCCGCAAGATGTCTCTGGAAGATCAACCGCTCAAGCGCTGGCGCG 4574  
 QY 435 CGGTAAACCGGTGATCGATCAAGGCAATTAATAAACCAGTGTGCAAAATACAA 494  
 DB 4573 GCAGTGTGTGTGACCGCGAAGCGTGGCGCGGAGACCGCGAGCTACTACGAGAC 4514  
 QY 495 AGCCGAAGCAAGCCGTTCAATATGGGATGATCTTCCGCGCGGATCTCAACATCAA 554  
 DB 4513 GGGCAAGCGCTGACCGCGAGACGCGCGCGGCTTCACTCAACAGTACCAACACC 4454  
 QY 555 ACTGCGTACTGCGCGCGCTGCGGTATCAATCTGCTACTATTCGCGCGCGAGGA 614  
 DB 4453 CGACAACTGAGGCGGCACTCAACACACCGGTCTGAGATCTTCAAGACGACCGAGG 4394  
 QY 615 CATTCCGGCCAAATCGGCGCAGAC 639  
 DB 4393 CAAGTTCGACTACTTGTGCGCGGC 4369

RESULT 11  
 US-09-902-540-5834

; Sequence 5834, Application US/09902540  
 ; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; PRIORITY FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 5834  
 ; LENGTH: 750  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 US-09-902-540-5834

Query Match 3.4%; Score 48; DB 4; Length 750;  
 Best Local Similarity 47.9%; Pred. No. 0.0086;  
 Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

144 GCTCACCGACATGCGCGCGTGGCGCGCGCAAAAGCTTCTTGAAGAGCAGAGG 203  
 DB 204 GCGCGGAGAGTTGCGCGAGGTGACGTGTCTGTCACAAACGCGGCGCTGGCGT 263

QY 204 CCTGTTCTGTCACTGGAAGCGGACCACTGGAAGTGTGTATGATAGGCTCGTGA 263  
 DB 264 GAGCTGGCGCAGTTCGCGCGCGCTGGAGACGTGGACAGATGTGTGAACCAACGTGA 323  
 QY 264 TGGCGAATGAGCGGCTGCAATGCTGGCGCGCGCGCTTAAAGCGCGCGCTTGGCT 323  
 DB 324 GGGCTCTCTGTACTGACGACGCGGTGCTGCGCGCGCATGTGGCGGACCGGAGCCA 383  
 QY 324 CGGCAACGAGCGGATATGAGGTGCGGTTCAGCATGAGCTTCAACGCAACCGCATTA 383  
 DB 384 CGTGTCAACATGAGGCTGCGGTGCGCGGAGGTGGCCCTAACCGCGCGGCAAGTGA 443  
 QY 384 GGTGTCAATGAATATGTCATCAATGAAGCGCAACATACCGCTGA 431  
 DB 444 CGCACCAAGCGCTTGTGACCACTTCAAGCTTCAACCTTCGCGCGGA 491

RESULT 12

US-09-902-540-1273/C

; Sequence 1273, Application US/09902540

; Patent No. 6833447

; GENERAL INFORMATION:

; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Wiegand, Roger C.  
 ; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 ; FILE REFERENCE: 38-10(15849)B  
 ; CURRENT APPLICATION NUMBER: US/09/902,540  
 ; PRIORITY FILING DATE: 2001-07-10  
 ; PRIOR APPLICATION NUMBER: 60/217,883  
 ; NUMBER OF SEQ ID NOS: 16825  
 ; SEQ ID NO 1273  
 ; LENGTH: 72704  
 ; TYPE: DNA  
 ; ORGANISM: Myxococcus xanthus  
 ; NAME/KEY: unsure  
 ; LOCATION: (1) (72704)  
 ; OTHER INFORMATION: unsure at all n locations  
 US-09-902-540-1273

Query Match 3.4%; Score 48; DB 4; Length 72704;  
 Best Local Similarity 47.9%; Pred. No. 0.046;  
 Matches 138; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

144 GCTCACCGACATGCGCGCGCTGCGCGCGCGCGCAAAAGCTTCTTGAAGAGCAGAGG 203  
 DB 70204 GCGCGGAGAGTTGCGCGAGGTGAGCGTGTGTCAACACGCGGCGCTGGCGGTT 70145  
 QY 204 CCTGTTCTGTCACTGGAAGCGGACCACTGGAAGTGTGTATGATAGGCTCGTGA 263  
 DB 70144 GAGCTGGCGCAGTTCGCGCGCGCTGGAGACGTGGACAGATGTGTGAACCAACGTGA 70085  
 QY 264 TGGCGAATGAGCGGCTGCAATGCTGGCGCGCGCGCTTAAAGCGCGCGCTTGGCT 323  
 DB 70084 GGGCTCTCTGTACTGACGACGCGGTGCTGCGCGCGCATGTGGCGGACCGGAGCCA 70025  
 QY 324 CGGCAACGAGCGGATATGAGGTGCGGTTCAGCATGAGCTTCAACGCAACCGCATTA 383  
 DB 70024 CGTGTCAACATGAGGCTGCGGTGCGCGGAGGTGCGCTTACCGCGGCGCAAGTGA 69965  
 QY 384 GGTGTCAATGAATATGTCATCAATGAAGCGCAACATACCGCTGA 431  
 DB 69964 CGCACCAAGCGCTTGTGACCACTTCAAGCTTCAACCTTCGCGCGGA 69917

RESULT 13

US-08-592-874-1/C

; Sequence 1, Application US/08592874  
 ; Patent No. 5854034

```

; GENERAL INFORMATION:
; APPLICANT: POLLOCK, THOMAS J.
; APPLICANT: YAMAZAKI, MOTOHIDE
; APPLICANT: THORNE, LINDA
; APPLICANT: MIKOLAJCZAK, MARCIA
; APPLICANT: ARMENTROUT, RICHARD W.
; TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCREASING
; TITLE OF INVENTION: POLYSACCHARIDE PRODUCTION
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JULES E. GOLDBERG
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/592,874
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/377,440
; FILING DATE: 24-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; US-08-592-874-1

Query Match      3.4%; Score 47.8; DB 2; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY      64 GTTTGGGGTTTGACGATGCGCCCGATGTCGGGGCAGTTGGCAAGCTGGAAAAAGGAAGAT 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18736 GTGCGGGGGCTGGCGATCAACACCGCCTGGCTGGCAATGCGCGCTGGAAGTGTGGCG 18677

QY      124 TTGAATTGCGCTTCATCAAGCTCACCGACATGGCGCGCTGGCGGTGCGCCGGA AAA 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18676 CCGACGCAAGGCGCGATCGCCCGATCGGCGAGACCAAGATGTCGATCCCCGAAAGC 18617

QY      184 GCGCTCTTGAGAGACGAGGGCCTGTTCTGTGCAACTGGAAGCCGAGGCCAATGGAAGTG 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18616 GGTATGTCGCGGCGATTCCTGTGGGGCGAGGGCGCAAGGTTTCAGAAAGGCCAGGTGCTG 18557

QY      244 GTGATGATAGGGTCTGATGAGCGCAACTGACGCGCTCCACATGCTGCGCGCGCGCGCG 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18556 ATCAAGCTGATGCCACATGTCGAGCGCAAGCCGCCGAGCGCGCGCGCTGCGCTGCTC 18497

QY      304 TTAGCGGCGACGCTTGCTTGGCACCAAGGCCGATATGAGGTGCGCTTCAGCATGGGC 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18496 AGCGCCCACTGATCCCGCGCAACAGCGCATCATGAGCGCTGAGCGGAAAGGGCG 18437

QY      364 TTCAACGGCAAGCGG 378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18436 TTCCGCTTCGCGCG 18422

```

```

RESULT 14
US-09-096-942-2/C
; Sequence 2, Application US/09096942
; Patent No. 6027925
; GENERAL INFORMATION:
; APPLICANT: Pollock, Thomas J
; APPLICANT: Mikolajczak, Marcia
; APPLICANT: Yamazaki, Motohide
; APPLICANT: Thorne, Linda
; APPLICANT: Armentrout, Richard W
; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
; TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
; FILE REFERENCE: seq 11st for appl filed from pro. appl
; CURRENT APPLICATION NUMBER: US/09/096,942
; CURRENT FILING DATE: 1998-06-12
; EARLIER APPLICATION NUMBER: 60/049,428
; EARLIER FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 28804
; TYPE: DNA
; ORGANISM: Sphingomonas sp. S88
; US-09-096-942-2

Query Match      3.4%; Score 47.8; DB 3; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY      64 GTTTGGGGTTTGACGATGCGCCCGATGTCGGGGCAGTTGGCAAGCTGGAAAAAGGAAGAT 123
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18736 GTGCGGGGGCTGGCGATCAACACCGCCTGGCTGGCAATGCGCGCTGGAAGTGTGGCG 18677

QY      124 TTGAATTGCGCTTCATCAAGCTCACCGACATGGCGCGCTGGCGGTGCGCCGGA AAA 183
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18676 CCGACGCAAGGCGCGATCGCCCGATCGGCGAGACCAAGATGTCGATCCCCGAAAGC 18617

QY      184 GCGCTCTTGAGAGACGAGGGCCTGTTCTGTGCAACTGGAAGCCGAGGCCAATGGAAGTG 243
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18616 GGTATGTCGCGGCGATTCCTGTGGGGCGAGGGCGCAAGGTTTCAGAAAGGCCAGGTGCTG 18557

QY      244 GTGATGATAGGGTCTGATGAGCGCAACTGACGCGCTCCACATGCTGCGCGCGCGCGCG 303
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18556 ATCAAGCTGATGCCACATGTCGAGCGCAAGCCGCCGAGCGCGCGCTGCGCTGCTC 18497

QY      304 TTAGCGGCGACGCTTGCTTGGCACCAAGGCCGATATGAGGTGCGCTTCAGCATGGGC 363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18496 AGCGCCCACTGATCCCGCGCAACAGCGCATCATGAGCGCTGAGCGGAAAGGGCG 18437

QY      364 TTCAACGGCAAGCGG 378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      18436 TTCCGCTTCGCGCG 18422

RESULT 15
US-09-096-867-2/C
; Sequence 2, Application US/09096867
; Patent No. 6030817
; GENERAL INFORMATION:
; APPLICANT: Pollock, Thomas J
; APPLICANT: Mikolajczak, Marcia
; APPLICANT: Yamazaki, Motohide
; APPLICANT: Thorne, Linda
; APPLICANT: Armentrout, Richard W
; TITLE OF INVENTION: Production of Xanthan Gum by Sphingomonas Bacteria
; TITLE OF INVENTION: Carrying Genes from Xanthomonas Campestris
; FILE REFERENCE: seq 11st for appl filed from pro. appl
; CURRENT APPLICATION NUMBER: US/09/096,867
; CURRENT FILING DATE: 1998-06-11
; EARLIER APPLICATION NUMBER: 60/049,428
; EARLIER FILING DATE: 1997-06-12
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 2
; LENGTH: 28804
; TYPE: DNA
; ORGANISM: Sphingomonas sp. S88
US-09-096-867-2

```

```

Query Match      3.4%; Score 47.8; DB 3; Length 28804;
Best Local Similarity 47.0%; Pred. No. 0.037;
Matches 148; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

```

```

QY 64 GTTTGGGGTTTGAAGATTGCGCCGCGATGTCGGGCGAATTGCAAGCTGGAAGAAAGAAAT 123
Db 18736 GTGGCGGGGGCTGGCGATCAACCGCCTGGCTGGCAATGGCCGCGTGAAGTGTGGCG 18677
QY 124 TTGAATTGGGCTTCATCAAGCTCACCGACATGGCGCCGCTGGCGGTGGCCCGAAGAA 183
Db 18676 CCGACCGCAGGGCGCGCATGCGCCCGATCGGCGAGACCAAGATCGTGCAATGCCCGAAGC 18617
QY 184 GGGTTCTTGGAGGACGAGGCGCTGTTGTCGCACTGGAAGCGGAGGCCAACTGGAAGTG 243
Db 18616 GGTATGTCGCGCGCATTTCTGTGGGCGAGGGCGAAGGTTGAGAAAGGCCAGGTGCTG 18557
QY 244 GTGATGATAGGGTCGTGAATGCGAATGAGCGGCTCGCACATGCTGGCGCGGCGCG 303
Db 18556 ATCAGCGCTGATCCCAACCAATGTGGGAGCCGGAAGCCGCGAGGCGCGCTGTGCTC 18497
QY 304 TTAGCGGCGACGCTTGGCTTGGCAACCAAGGCCGATATCGAGGTGCCGTTCAAGCATGGGC 363
Db 18496 AGCGCCAGCTCGATCCGCGCGCAACCAAGCGATCATCGACGCTGGACGGAAGGGC 18437
QY 364 TTCAAGCGCAAGCGG 378
Db 18436 TTCCGCTTCGCGCG 18422

```

```

Search completed: September 16, 2005, 02:48:44
Job time : 274 secs

```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: September 15, 2005, 22:21:25 ; Search time 928 Seconds  
(without alignments)  
9993.857 Million cell updates/sec

Title: US-10-689-200-1

Perfect score: 1392  
Sequence: 1 atgaataaccatcatatagatc.....ccggcgcaagtcgtcgat 1392

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7389322 seqs, 333128559 residues

Total number of hits satisfying chosen parameters: 14778644

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Published Applications NA:\*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14:	/cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17:	/cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
18:	/cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
19:	/cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
20:	/cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
21:	/cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
22:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
23:	/cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*
24:	/cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
25:	/cgn2_6/ptodata/1/pubpna/US6_NEW_PUB.seq:*
26:	/cgn2_6/ptodata/1/pubpna/US6_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	
1	1392	100.0	1392	19	US-10-689-200-1	
2	143	10.3	551	19	US-10-689-200-3	
3	71.8	1209	17	US-10-282-122A-33791	Sequence 3, Appli	
4	62.2	4.5	1173	17	US-10-282-122A-11998	Sequence 11998, A
5	59.2	4.3	576	17	US-10-283-122A-11492	Sequence 11492, A
6	58	4.2	1200	17	US-10-283-122A-30186	Sequence 30186, A
7	58	4.2	1200	18	US-10-389-647-149	Sequence 149, App

8	57	4.1	1116	17	US-10-282-122A-32179	Sequence 32179, A
9	49.4	3.5	1512	17	US-10-369-493-42267	Sequence 42267, A
10	48.4	3.5	773	17	US-10-282-122A-8343	Sequence 8343, Ap
11	48	3.4	744	17	US-10-369-493-43243	Sequence 43243, A
12	47.8	3.4	1017	19	US-10-437-963-85118	Sequence 85118, A
13	47.8	3.4	7656	17	US-10-282-122A-25488	Sequence 25488, A
14	46.4	3.3	3048	15	US-10-156-761-3146	Sequence 3146, Ap
15	46.4	3.3	9025608	15	US-10-156-761-1	Sequence 1, Appli
16	46.2	3.3	1440	19	US-10-437-963-82648	Sequence 82648, A
17	45.6	3.3	1173	17	US-10-282-122A-13773	Sequence 13773, A
18	45.4	3.3	1396	19	US-10-767-701-13423	Sequence 13423, A
19	44.4	3.2	930	19	US-10-437-963-93713	Sequence 93713, A
20	44.4	3.2	1365	19	US-10-437-963-56913	Sequence 56913, A
21	44	3.2	786	19	US-10-437-963-13840	Sequence 13840, A
22	43.4	3.1	1188	17	US-10-282-122A-13623	Sequence 13623, A
23	43.4	3.1	5577	17	US-10-369-493-43212	Sequence 43212, A
24	43.2	3.1	1155	19	US-10-437-963-67845	Sequence 67845, A
25	43.2	3.1	2110	19	US-10-437-963-10628	Sequence 10628, A
26	43	3.1	1305	17	US-10-282-122A-11478	Sequence 11478, A
27	43	3.1	1818	17	US-10-369-493-27231	Sequence 27231, A
28	43	3.1	2021	19	US-10-437-963-49750	Sequence 49750, A
29	42.8	3.1	956	18	US-10-425-114-24149	Sequence 24149, A
30	42.8	3.1	984	20	US-10-425-115-118414	Sequence 118414, A
31	42.8	3.1	993	19	US-10-437-963-93714	Sequence 93714, A
32	42.6	3.1	2213	21	US-10-772-636-27	Sequence 27, Appli
33	42.6	3.1	2231	17	US-10-282-122A-15175	Sequence 15175, A
34	42.4	3.0	1122	15	US-10-156-761-4985	Sequence 4985, Ap
35	42.4	3.0	1305	17	US-10-282-122A-12878	Sequence 12878, A
36	42.4	3.0	1470	9	US-09-815-242-7855	Sequence 7855, Ap
37	42.4	3.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
38	42.2	3.0	375	18	US-10-425-114-4691	Sequence 4691, Ap
39	42.2	3.0	501	18	US-10-425-114-319	Sequence 319, Appl
40	42.2	3.0	726	18	US-10-425-114-58	Sequence 58, Appl
41	42.2	3.0	1525	17	US-10-369-493-27882	Sequence 27882, A
42	42.2	3.0	1866	20	US-10-425-115-74324	Sequence 74324, A
43	42	3.0	536	17	US-10-338-110-119	Sequence 119, App
44	41.6	3.0	900	17	US-10-369-493-41628	Sequence 41628, A
45	41.6	3.0	1271	19	US-10-437-963-38282	Sequence 38282, A

## ALIGNMENTS

RESULT 1  
US-10-689-200-1  
Sequence 1, Application US/10689200  
Publication No. US20040126848A1  
GENERAL INFORMATION:  
APPLICANT: Dicostino, Deana J.  
APPLICANT: Ni, Hao  
APPLICANT: Ye, Rick  
APPLICANT: Picataggio, Stephen  
APPLICANT: Wang, Tao  
APPLICANT: Seip, John E.  
TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN CL METABOLIZING BACTERIA  
FILE REFERENCE: CL1747 US NA  
CURRENT APPLICATION NUMBER: US/10/689,200  
CURRENT FILING DATE: 2003-10-20  
PRIOR APPLICATION NUMBER: 60/419,872  
PRIOR FILING DATE: 2002-10-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 1  
LENGTH: 1392  
TYPE: DNA  
ORGANISM: Methylobionas sp. 16a  
US-10-689-200-1  
Query Match 100.0%; Score 1392; DB 19; Length 1392;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 ATGAATAACCATCATATAGATCGAGCTCGAAGAAATTGTTATTCAGCTTATCGGCTTCGCTA 60

```

Db      1 ATGAAAACCATCATATGATCGAGCTCGAAGAAATGTATATGACGTTATCGGCTTGCTA 60
Qy      61 GCCGTTGGGGTTTGACGATTTGCGCCGAGTTCGGGGCAGTTGGAGCAAGCTGAGAAAAGAA 120
Db      61 GCCGTTGGGGTTTGACGATTTGCGCCGAGTTCGGGGCAGTTGGAGCAAGCTGAGAAAAGAA 120
Qy      121 GATTGAAATTCGGCTTCAATCAAGCTCACCGCATGCGCCGCTGGCGGTGGCCGCGAA 180
Db      121 GATTGAAATTCGGCTTCAATCAAGCTCACCGCATGCGCCGCTGGCGGTGGCCGCGAA 180
Qy      181 AAAGGCTTCTTGAGAGACGAGGGCTGTTGTCGCACTGGAAGCGCAGCCCAACTGGAAG 240
Db      181 AAAGGCTTCTTGAGAGACGAGGGCTGTTGTCGCACTGGAAGCGCAGCCCAACTGGAAG 240
Qy      241 GGGGTATGATGAGGTGCGTGAATGGCGAATGGCGATCGGCTGCGGCGCGCGCG 300
Db      241 GGGGTATGATGAGGTGCGTGAATGGCGAATGGCGATCGGCTGCGGCGCGCGCG 300
Qy      301 CGGTTAGCGGCGACGCTTGAGCTTTCGACCAAGGCGATTCGAGGTGCGGTTCAAGCAT 360
Db      301 CGGTTAGCGGCGACGCTTGAGCTTTCGACCAAGGCGATTCGAGGTGCGGTTCAAGCAT 360
Qy      361 GGGTTCAACGGCAACGCGATTCAGGTGTCGAATGAATTCGGCATCAGATGAAGCCGAAC 420
Db      361 GGGTTCAACGGCAACGCGATTCAGGTGTCGAATGAATTCGGCATCAGATGAAGCCGAAC 420
Qy      421 ATACCGCTGGAAGGGCGGTAAACCGGTGACATCGGATGAAGGCGATTAATTAACCGGTC 480
Db      421 ATACCGCTGGAAGGGCGGTAAACCGGTGACATCGGATGAAGGCGATTAATTAACCGGTC 480
Qy      481 GTGAAAAAATACAAAGCCGAGGCAAGCGGTTCAATATGCGATGACCTTCCGCGCGGA 540
Db      481 GTGAAAAAATACAAAGCCGAGGCAAGCGGTTCAATATGCGATGACCTTCCGCGCGGA 540
Qy      541 TCTCACAACATCAAACTGCGTTACTGCTGCGGCGGTGCGGCTGATCAATCTGCTACTAT 600
Db      541 TCTCACAACATCAAACTGCGTTACTGCTGCGGCGGTGCGGCTGATCAATCTGCTACTAT 600
Qy      601 TGGCGCGCGCAGAGACATTCGCGGCAAAATCGGCGGAGAGCGCTGTTGTCGGTGACCCG 660
Db      601 TGGCGCGCGCAGAGACATTCGCGGCAAAATCGGCGGAGAGCGCTGTTGTCGGTGACCCG 660
Qy      661 CGCGCGCAAAATCGCGTCAACGCTGGAAGCGGCAACCATTTTGGTTATTCGTCGCGAG 720
Db      661 CGCGCGCAAAATCGCGTCAACGCTGGAAGCGGCAACCATTTTGGTTATTCGTCGCGAG 720
Qy      721 CCGTGAACCAAGCGGCGGTTTCAAGGGGATGAGCGGTGCGGTGATCACCGATGAAGAA 780
Db      721 CCGTGAACCAAGCGGCGGTTTCAAGGGGATGAGCGGTGCGGTGATCACCGATGAAGAA 780
Qy      781 CTCTGGAAGGACACGCGCGGAAAAAGTCTTCGGGTGACCAAAATTCGGCGGAGAAATTC 840
Db      781 CTCTGGAAGGACACGCGCGGAAAAAGTCTTCGGGTGACCAAAATTCGGCGGAGAAATTC 840
Qy      841 CCCAACACCTATCTGCGGTGACCAAGGCGCTGATTCGGGCGCGCATCTGCTGAGCGCC 900
Db      841 CCCAACACCTATCTGCGGTGACCAAGGCGCTGATTCGGGCGCGCATCTGCTGAGCGCC 900
Qy      901 GACATATACAAAGACCGCAAGGACCATGAAATGCTGCGCAAAAAACAATACGTCGAT 960
Db      901 GACATATACAAAGACCGCAAGGACCATGAAATGCTGCGCAAAAAACAATACGTCGAT 960
Qy      961 GCGTGAAGTGAAGTGTGCGGCTGAGCAAGACGCGCATTCGAAATGACAAAGACAT 1020
Db      961 GCGTGAAGTGAAGTGTGCGGCTGAGCAAGACGCGCATTCGAAATGACAAAGACAT 1020
Qy      1021 AAACGCGCGCTACCGGACTTCAACCTTCTTTCGCGCAGCGCGCAGCTATCCGCTCTAC 1080
Db      1021 AAACGCGCGCTACCGGACTTCAACCTTCTTTCGCGCAGCGCGCAGCTATCCGCTCTAC 1080
Qy      1081 AGCAGTGCAGTCTGATCTGAACCCAGCTGAAGCGCTGGGGCATGATCAATGAATTCAA 1140
Db      1081 AGCAGTGCAGTCTGATCTGAACCCAGCTGAAGCGCTGGGGCATGATCAATGAATTCAA 1140

```

```

Db      1081 AGCAGTGCAGTCTGATCTGAACCCAGCTGAAGCGCTGGGGCATGATCAATGAATTCAA 1140
Qy      1141 CGGACCAACTGATATCTGATATCCGCGCAAGACGCTACCGCGCGGACATCTATCTCGCC 1200
Db      1141 CGGACCAACTGATATCTGATATCCGCGCAAGACGCTACCGCGCGGACATCTATCTCGCC 1200
Qy      1201 GCGGCGCAAGGAACTGATCGCAGAAAGCAAGGCGCAAGGCGAAGACTTCCTGCGCATAC 1260
Db      1201 GCGGCGCAAGGAACTGATCGCAGAAAGCAAGGCGCAAGGCGAAGACTTCCTGCGCATAC 1260
Qy      1261 TCGATCAAGCGGTCGCAAGATTTCTTATCCGAAAGTGCCTGATGACCAAGACCC 1320
Db      1261 TCGATCAAGCGGTCGCAAGATTTCTTATCCGAAAGTGCCTGATGACCAAGACCC 1320
Qy      1321 AACGATTAACCTGCGCAAGTTTGCATAGGTCTGAAGGCAAGCAACCGTAGCCGCGGC 1380
Db      1321 AACGATTAACCTGCGCAAGTTTGCATAGGTCTGAAGGCAAGCAACCGTAGCCGCGGC 1380
Qy      1381 AAGTCTGTGAT 1392
Db      1381 AAGTCTGTGAT 1392

```

```

RESULT 2
US-10-689-200-3
; Sequence 3, Application US/10689200
; Publication No. US20040126848A1
; GENERAL INFORMATION:
; APPLICANT: Dicosimo, Deana J.
; APPLICANT: Ni, Hao
; APPLICANT: Ye, Rick
; APPLICANT: Picataggio, Stephen
; APPLICANT: Wang, Tao
; APPLICANT: Seip, John E.
; TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN CL METABOLIZING BACTERIA
; FILE REFERENCE: CL1747 US NA
; CURRENT APPLICATION NUMBER: US/10/689,200
; PRIOR FILING DATE: 2003-10-20
; PRIOR APPLICATION NUMBER: 60/419,872
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 3
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Methylobionas sp. 16a
; NAME/KEY: misc_feature
; LOCATION: (409)..(411)
; OTHER INFORMATION: ATG start site of nrfA gene
US-10-689-200-3

```

```

Query Match      10.3%; Score 143; DB 19; Length 551;
Best Local Similarity 100.0%; Pred. No. 7.7e-34;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 ATGAAAACCATCATATGATCGAGCTCGAAGAAATGTATATGACGTTATCGGCTTGCTA 60
Db      409 ATGAAAACCATCATATGATCGAGCTCGAAGAAATGTATATGACGTTATCGGCTTGCTA 468
Qy      61 GCCGTTGGGGTTTGACGATTTGCGCCGAGTTCGGGGCAGTTGGAGCAAGCTGAGAAAAGAA 120
Db      469 GCCGTTGGGGTTTGACGATTTGCGCCGAGTTCGGGGCAGTTGGAGCAAGCTGAGAAAAGAA 528
Qy      121 GATTGAAATTCGGCTTCAATCAAGCTCACCGCATGCGCCGCTGGCGGTGGCCGCGAA 180
Db      529 GATTGAAATTCGGCTTCAATCAAGCTCACCGCATGCGCCGCTGGCGGTGGCCGCGAA 584

```

```

RESULT 3
US-10-282-122A-33791
; Sequence 33791, Application US/10282122A
; Publication No. US20040029129A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33791

Query Match          5.2%; Score 71.8; DB 17; Length 1209;
Best Local Similarity 53.8%; Pred. No. 2,1e-11;
Matches 148; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 102 TGGCAAGCTGGAAGAGAGATTGAAATTGGCTTCATCAAGCTCAGCGACATGGCGCC 161
DB 45 TGAGCGCCCGCGAATGAAACAGCTGACGTGGGCTTCATGSCACTGACGACTGGCGCC 104
QY 162 GCTGGGGGTGGCGCGGAAAGAGCTTCTTGAGAGAGAGGCGCTGTTCTGGAACCTGGA 221
DB 105 GCTGGGTGTGCTGCCACCGAGGGGTTGCCAGCCTTACGGGCTGAGCCTGGAACCTGGA 164
QY 222 AGCGAGCGCCACTGGAAGGTGTGATGATAGGGGTGTGATGAGCGAACTGGACGGCTTC 281
DB 165 AGCGCAGAGCTCTGCGGCGCGGGCTGCGGTGACAGGCTGTCAAGGGAGCTTCAAGCGCG 224
QY 282 GCAATGCTGGCGCGCGCGCGCTTGAAGGCGCAAGGCTTGGCGCACCAAGCCGATAT 341
DB 225 ACATGAGCTGTATGATTGATCTATGCGCTGAGAGCTGGGCGATAGCGCGCGCGCGAG 284
QY 342 CGAGGTGCGGTTCAGATGAGGTGCTTCAACGCGCAAGC 376
DB 285 CGACATGGCCATCTCATGGGCTCAACCAAGAGC 319

RESULT 4
US-10-282-122A-11998
```

```

; Sequence 11998, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11998
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-282-122A-11998

Query Match          4.5%; Score 62.2; DB 17; Length 1173;
Best Local Similarity 58.4%; Pred. No. 2,1e-08;
Matches 128; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY 131 TCGGCTTCACTCAAGCTCAGCAGATGGCGCGCTGGCGCGCGGAAAGGCTTCT 190
DB 116 TCGGCTATGTCCGATCAGCAGCGCGCCCGCTGTGTGCTGCCACACACAGCGCTATT 175
QY 191 TCGAGAGCAGAGGCGCTGT--TCGTGCAACTGGAAGCGCAAGCTGGAAGGTGTGA 247
DB 176 TCGCGTCCGAAGGGCTGAGCGGTGAGCAACGAAAGCTGTGGGTAGCTGGGCGCAGCTGG 235
QY 248 TGAATAGGTCGTGAATGCGGAAGTGAAGGCTTCGACATGCTGGGCGCGCGCGCTTGA 307
DB 236 TCGAGGCGTTCTGTCGCGCGCAGGTCAACGTGTGCACTGCTGTGCGCGATGACGCTGT 295
QY 308 CGGCGAGCGTTGGCTTGGCAGCAGCGCCGATATCGAGG 346
DB 296 GGGCGCGCTACGCGCAGCCGCGCGCGCGCGCGAAGTGTGG 334

RESULT 5
US-10-282-122A-11492
; Sequence 11492, Application US/10282122A
; Publication No. US20040029129A1
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 11492
/ LENGTH: 576
/ TYPE: DNA
/ ORGANISM: Burkholderia cepacia
US-10-282-122A-11492

Query Match      4.3%; Score 59.2; DB 17; Length 576;
Best Local Similarity 57.9%; Pred. No. 1.4e-07;
Matches 125; Conservative 0; Mismatches 88; Indels 3; Gaps 1;

QY      131 TCGGCTTCATCAAGCTCACGACATGCGCGCTGGCGGTGGCGCGCAAAAAGGCTTCT 190
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      116 TCGGCTATCTGCCGATACGACGCGCGCGCTGCTGTGCTCCACACACGCTATT 175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      191 TCGAGACGAGGCGCTGT--TCGTGCACTGGAGCGCAGCCAACTGGAAAGTGATGA 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      176 TCGGCTCCGAAGGCGTGAAGGTGAGGACCAAGCGCTGCGCACTGGCGCACTG 235
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      248 TGGATAGGCTGTGAATGGGAACTGAGCGGCTCGCATGCTGGCGCGCGCGCTTGA 307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      236 TCGAGGCGTTCCCTCGCGGCGCAGGTCAACGTCTGTGACCTGTGCGCGATGACGCTGT 295
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      308 CGGCGAGCGTTGGCTTCGGACACCAAGGCGGATATCG 343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      296 GGGCAGCTAGCGGACGCGCGCGCGCGCGCAAGGTGG 331
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-10-282-122A-30186
/ Sequence 30186, Application US/10282122A
/ Publication No. US2004002919A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangu
```

```
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kari
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 30186
/ LENGTH: 1200
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30186

Query Match      4.2%; Score 58; DB 17; Length 1200;
Best Local Similarity 56.2%; Pred. No. 4.5e-07;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

QY      131 TCGGCTTCATCAAGCTCACGACATGCGCGCTGGCGCGCGCAAAAAGGCTTCT 190
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      137 TCGGCTACCTGCCGATCACGACGCGCAGCCCTTGTCTGTGCTCCATGCGACGCTCT 196
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      191 TCGAGACGAGGCGCTGTTCGTGCACTGGAAAGCGCAGCC--AACTGAAAGTGATGA 247
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      197 TCGAGGCGGAGGCGCTCCAGCGCGCAACGCGCGGTGCTGTGCGTACTGGCGCAGTGA 256
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      248 TGGATAGGCTGTGAATGGGAACTGAGCGGCTCGCATGCTGGCGCGCGCGCGCTTGA 307
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      257 TCGAGGCGTTCACTCCCGCGCAGTCAACGTATCACCCTGTGTGCGCGATGACGCTGT 316
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      308 CGGCGAGCGTTGGCTTCGGACCAAGGCGGATATGAGAGTGCGCTTACAGCATGGGCTTCA 367
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      317 GGGGCC-----CGTACGGCAGCAAGTGCGCGCGCAAGTGATGATGCGCATGCGTGC 370
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY      368 ACGGCAAGCGGATTACGGGTGTCAATGAATC 399
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      371 GCGGCTCGGGCTCACGGTGGCTCCGAGATC 402
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-10-389-647-149
/ Sequence 149, Application US/10389647
/ Publication No. US2004003549A1
```

```

; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Cami
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 149
; LENGTH: 1200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-149

```

```

Query Match          4.2%; Score 58; DB 18; Length 1200;
Best Local Similarity 56.2%; Pred. No. 4; Se-07;
Matches 153; Conservative 0; Mismatches 110; Indels 9; Gaps 2;

```

```

QY 131 TCGGCTTCATCAAGCTCAGCAGCATGGCGCGCTGGCGGTGGCCGCGAAAGGCTTCT 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 137 TCGGCTTACTGCGCATACCGACGCCACGCTTGTGTGGCCCATGCCAACGGCTGT 196
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 TCGAGACGAGAGCGCTTGTGCACTGGAAGCGCAGGCC---AACTGGAAGTGGTGA 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 197 TCGAGGCCGAGGGGATCCAGAGCGAGCGCGGTGCTGTTCGTAAGTGGGCCAGTGA 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 TGGATAGGCTGCGAATGCGGAAGTGAAGCGCTTCGACATGCTGGCGCGCGCCGTAG 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 257 TCGAGGCGTCACTCCGCGCAGGTCAACGTCACTGCTGTGCGCCGATGACCGTCT 316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 CGGCGACGCTTGGCTTGGGACCAAGGCCGATATCGAGGTGGCGCTTCAGATGGGCTTCA 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 317 GGGCC-----CGTACGGGAGCAAGGTGCCGGCCAAAGGTGGGCTTGAACCAAGCTCG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 ACGCAACGCCATTACGGTGTCCATGCAATGAATC 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 371 GCGGCTCGGGGCTCACGGTGTCCGAGATC 402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 8
US-10-282-122A-32179
; Sequence 32179, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: Elittra.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335

```

```

; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32179
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-32179

```

```

Query Match          4.1%; Score 57; DB 17; Length 1116;
Best Local Similarity 54.1%; Pred. No. 8; Se-07;
Matches 139; Conservative 0; Mismatches 115; Indels 3; Gaps 1;

```

```

QY 131 TCGGCTTCATCAAGCTCACCAGCATGGCGCGCTGGCGGTGGCCGCGAAAGGCTTCT 190
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 56 TCGGCTTACTGCGCATACCGACGCCACGCTTGTGTGGCCCATGCCAACGGCTGT 115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 TCGAGACGAGAGCCCT---GTTGTGCACTGGAAGCGCAGGCCAAGTGAAGTGTGA 247
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 116 TCGAGGCCGAGGGGATCCAGAGCGAGCGCGGTGCTGTGCGCACTGGGCCAGGTGA 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 TGGATAGGCTGCGAATGCGGAAGTGAAGCGCTTCGACATGCTGGCGCGCGCCGTAG 307
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 176 TCGAGGATTCATCTCCGCGCAGGTCAATGTATCACTGCTTTCGCGATACCGTAT 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 CGGCGACGCTTGGCTTGGGACCAAGGCCGATATCGAGGTGGCGCTTCAGATGGGCTTCA 367
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 236 GGGCGCGCTATGGCAGCAGAGTGCAGCCAAAGGTGGTGCCTGGAACCAAGTGGCGGTT 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 368 ACGCAACGCCATTACG 384
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 296 CGGCTCTGACTGTTCG 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 9
US-10-369-493-42267
; Sequence 42267, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42267
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42267

```

```

Query Match          3.5%; Score 49.4; DB 17; Length 1512;
Best Local Similarity 47.8%; Pred. No. 0.00024;

```



```

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25486
; LENGTH: 7656
; TYPE: DNA

```

Query Match	3.4%	Score 47.8	DB 17	Length 7656
Best Local Similarity	54.2%	Pred 0.0015		
Matches 97, Conservative	0	Mismatches 82	Indels 0	Gaps 0

Qy	159	GCCGCTGACGGTGGCGCGCCGAAAAAGGCTTCTTGAGAGACGAGGGACCTGTTGCGCACT	218
Db	1929	GACCAATGTTCTGGCCCGCGCCGTCGAGGCGTACCGCGACGACGTCCGCGTGTGCTGCA	1988
Qy	219	GGAAGCGCAGGCCCAACTGGAAGGTGTGATGATGATGAGGTCGTGATGGCGAACTGACGCG	278
Db	1989	GGCGCTGTGGAACGCCACCCGATGCTGCGGCTGCGGGTCGAGGACGACGGCGCCGCGCG	2044
Qy	279	CTGCGACATGCTGAGCGCGCGCGCGCTTGAGCGGCGAGCGTTGGCTTTCGACACCAAGCGCG	337
Db	2049	CTGGAACTGTGACAGCGCGCGAGCCCGGCTCGGTGCGGCGCGCCGACCTGCTGCAGTCCG	2107

RESULT 14  
US-10-156-761-3146  
; Sequence 3146, Application US/10156761  
; Publication No. US20030119018A1

```

RESULT 14
US-10-156-761-3146
; Sequence 3146, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMOBA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HOKIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3146
; LENGTH: 3048
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3048)
US-10-156-761-3146

```



Query Match 3.3%; Score 46.4; DB 15; Length 3048;  
Best Local Similarity 50.0%; Pred. No. 0.0028;  
Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Search completed: September 16, 2005, 05:01:18  
Job time : 949 secs

QY 69 GGGTTGACGATTGCGCCCGATGTGCGGCGAGTTGCGAAGCTGAAAAGAGATTGAA 128  
DB 645 GGTGCTGACGACCGCCCTGTAACGTGCTGCGCCCTACACCGCCAGAGAGACATCCC 704  
QY 129 ATTGGCTTCATCAAGTCAACGACATGAGCGCGCTGCGCGTGGCCCGGAAAAAGCTT 188  
DB 705 GCTCGGATCCCATGCTCGGCGCGCACGAGCCGAACTCGAGACGTCGTGCGCTGTT 764  
QY 189 CTTGAGAGACGAGGCGCTGTTCTGCACTGGAAGCGCAGCCCACTGGAAGGTGAT 248  
DB 765 CATCAACATGCTGTGTGTGCGCACCGATCTGTCCGGGAGACCCACTTCCGCGAGCTGAT 824  
QY 249 GGTAGGGTGTGTAATGCGCACTGGAAGGCTGCACTGCACTGCGCGCGCGCG 300  
DB 825 CGAGCGGGTGCAGGACGCCAATGACGTGTACGACACACGAGACGTGCGCG 876

## RESULT 15

US-10-156-761-1/c  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 3.3%; Score 46.4; DB 15; Length 9025608;  
Best Local Similarity 50.0%; Pred. No. 0.076;  
Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 69 GGGTTGACGATTGCGCCCGATGTGCGGCGAGTTGCGAAGCTGAAAAGAGATTGAA 128  
DB 3936846 GGTGCTGACGACCGCCCTGTAACGTGCTGCGCCCTACACCGCCAGAGAGACATCCC 3936787  
QY 129 ATTGGCTTCATCAAGTCAACGACATGAGCGCGCTGCGCGTGGCCCGGAAAAAGCTT 188  
DB 3936786 GCTCGGATCCCATGCTCGGCGCGCACGAGCCGAACTCGAGAGATGCTGCGCTGTT 3936727  
QY 189 CTTGAGAGACGAGGCGCTGTTCTGCACTGGAAGCGCAGCCCACTGGAAGGTGAT 248  
DB 3936726 CATCAACATGCTGTGTGTGCGCACCGATCTGTCCGGGAGACCCCACTTCCGCGAGCTGAT 3936667  
QY 249 GGTAGGGTGTGTAATGCGCACTGGAAGGCTGCACTGCACTGCGCGCGCGCG 300  
DB 3936666 CGAGCGGGTGCAGGACGCCAATGACGTGTACGACACGAGACGTGCGCG 3936615

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 00:58:27 ; Search time 42 Seconds

(without alignment)  
1062.966 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443

Sequence: 1 MTTIRSSSKLLTLTSLA.....AKFATGKQTVAGKQVD 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 79: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	651	26.6	670	2	S77387
2	627.5	25.7	430	2	AC3037
3	627.5	25.7	471	2	H98248
4	616.5	25.2	441	2	E95928
5	611.5	25.0	431	2	AD3037
6	611.5	25.0	431	2	G98248
7	610.5	25.0	440	2	AG1882
8	587	24.0	657	2	A11882
9	576	23.6	430	2	C95300
10	572	23.4	443	2	J02134
11	549.5	22.5	430	2	G87324
12	545.5	22.3	442	2	S58738
13	538.5	22.0	667	2	AH2165
14	534.5	21.9	667	2	S75959
15	525	21.5	446	2	S77389
16	502.5	20.6	458	2	AF2165
17	494	20.2	402	2	F83422
18	476	19.5	659	2	S30893
19	473.5	19.4	452	2	S75956
20	470.5	19.3	427	2	AF2222
21	444	18.2	450	2	A44751
22	398.5	16.3	392	2	S52248
23	390.5	16.0	426	2	AF2663
24	390.5	16.0	426	2	E97445
25	355	14.5	625	2	S58789
26	249	10.2	329	2	B55581
27	240	9.8	399	2	A83354
28	125	5.1	337	2	B75370
29	124.5	5.1	333	2	AB3080

30	124.5	5.1	333	2	G98206	hypothetical prote
31	116.5	4.8	468	2	A83359	hypothetical prote
32	114	4.7	325	2	H75414	probable ABC trans
33	114	4.7	336	2	A83801	ABC transporter (b
34	111	4.5	310	2	AC3238	hypothetical prote
35	109.5	4.5	300	2	H69260	thiamin biosynthes
36	106.5	4.4	675	2	D48492	kpac protein - Bac
37	106	4.3	483	2	G86811	sugar ABC transp
38	106	4.3	1273	2	T34558	hypothetical prote
39	105.5	4.3	339	2	E95962	probable taurine u
40	105.5	4.3	566	2	A40589	cellulase (EC 3.2.
41	104.5	4.3	1135	2	T14699	DNA polymerase III
42	104.5	4.3	1174	2	T15021	probable DNA poly
43	103	4.2	897	2	AC2001	hypothetical prote
44	102	4.2	301	2	B95947	phosphate uptake A
45	102	4.2	1291	2	I40631	non-proteolytic bo

#### ALIGNMENTS

##### RESULT 1

S77387 nitrate transport protein C-2 - *Synechocystis* sp. (strain PCC 6803)

N/Alternate names: protein e111452

C/Species: *Synechocystis* sp.

A/Variety: PCC 6803

C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 16-Aug-2004

C/Accession: S77387

R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

8.

A/Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S77387

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-670 <KAN>

A/Cross-references: UNIPROT:P73450; EMBL:D90906; GB:AB001339; NID:G1652492; PIDN:BAA174;

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Genetics:

A/gene: nrtC-2

C/Superfamily: ATP-binding cassette homology

C/Keywords: ATP; nucleotide binding; P-loop

F/25-315/Domain: ATP-binding cassette homology <ABC>

F/42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 26.6%; Score 651; DB 2; Length 670;  
Best Local Similarity 36.3%; Pred. No. 9.7e-40;

Matches 158; Conservative 68; Mismatches 153; Indels 56; Gaps 13;

QY	28	APDVGAVGK-----LEKEDLKFGFKLTDMAPLAFAEKGFEEDEGL-FVQLEAOA	77
DB	260	AKKVGAVSQFAPAMGANGKELNLDGRIPLDCAPLVAAKKGKFGQKGLGQVNLVXEP	319
QY	78	NKKVVMRDVNVGELDGSFHLAPAPLAASVGFSTKADIEV--PFSKGFNGAATVTSNEIWH	135
DB	320	SMOATADGIRREKRLDOAQVAGMPPLATLGMGSKTPLPVATVMVNSRNAGATLTSKFAE	379
QY	136	QKKPNIPLEGKGFVHPIKADYIKPVVEKKYKAGKGFNNAMTTPAGSHNFKLAVYLAAGGI	195
DB	380	-----AGVTLDELR-----LKAETPDQVSTLGMVHPASQNLTLVYWLASGSI	424
QY	196	NGGVYSPQDIDSGQIGADALISVPPQPMSTLEAGTTFGVCVGEPMNQOAVFGIGIVPV	255
DB	425	D-----PDQDIN-----LMRLPPQMVNSLEAGNIDGFCVGEPMNSYAAKQNLGYVI	471
QY	256	ITDEELMKDTPKPKVGVGTQWAKETPNYTLAVTKLIRAPALWLDADNNKRRKAEITMLAQ	315
DB	472	ATVDLIDWNGHPRKVLVGMREEVNKKYPATHTLALVKKALLEACEY--CDDRHRROEIIDYAL	529
QY	316	KQYVADVEVLAASNNGTETFEKQ--DKRALPFPNFFRRGASVPSSAVWVLTQLRRNG	374

Db 530 P O Y V T S T E Y I S P G F L T E Y D G N D A E A E M L D F N G F Y V K Q S N Y P S R S S G L M T L Q L A R W G 589

Qy 375 M I N E K P N M W Y L D T K N Y R P D I Y L A A K E L V A E K A A E D P P A T S I K P S O N F F I D K V P 434

Db 560 Y I D -- F P K W - V E I E R V R R P D L F S E A C R H L --- C M P D L E G D H H V S L ----- F D C M V 636

Qy 435 F D A N K P N D I A K A F A I 449

Db 637 F T P N D P L G Y I K R F T I 651

## RESULT 2

Hypothetical protein nr1a [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
 C:Accession: AC3037  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McTiell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tso, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 seer, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AC3037  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-430 <KUR>  
 A:Cross-references: UNIPROT:O8U925; GB:AE008669; PDB:1AAL44713.1; PID:g17424344; GSPDB:  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: nr1a  
 A:Map position: linear chromosome

Query Match	25.7%	Score 627.5	DB 2	Length 430
Best Local Similarity	35.5%	Pred. No. 2.7e-38		
Matches 166	Conservative	70	Mismatches 171	Indels 61
				Gaps 19

```

QY      1 KKTIIRSS-SKKLL--TLSASLAWGSLIAPD--VGA VKLEKEDLKRFGFKYLTDMAPL 55
Db      1 MKCIFSGVSRRTILKTTATTAALVTAVRTAFSGAFPAATAPEVNGALGFALTDAAPL 60
QY      56 AVAAEKGFEEDEGL-FVQLEAQAANKVMDRVNAG---ELDGSHTLAPAPLAASVGRGT 110
Db      61 IIAEKGFAPAKHGMEVEVLKQASWGAIRNDLVLGASNGIDGAILILPMPYIMHTGKVT 120
QY      111 KADIEVPS---MGFNGNATTVSNEIWHQMKENIPLBEGKRVPIKADYLKPVVEKTKA 166
Db      121 QNNPVPMTILARLNLDSQGISVAKG---YARTQQLDASK-----LKAFAEKKKA 168
QY      167 EGPENPMAMTEPPAGSHNIKLRYWYLAAGINGVYSPPODISQIGADALLSTVPPQWMS 226
Db      169 DQKEIKAMTTPPGGTHDMIRYWLAAAGSID-----PDKDVS-----TIVPEPPQWMA 215
QY      227 TLEAGTIFGVCYGEWBNOQAVFKGIGVPIITDEELMKDTPKRFVSGVTROMAEKYPTNYLA 286
Db      216 NMKVGNMNVPCVGEWBNQOLVNOGIGFTACTGELMKGHPEKALQMRADWEKPNPATKA 275
QY      287 VTKALIRAPAILMDADNNKQRKEALEMLAKOKVQGVADVEVLAAASNGTTEYEEKDKRALPD 346
Db      276 LIMAAMEAQW--CEEMANKKEBSTILGKROMFNVPVDVYGRLLKGNINY--GNGRVYLE- 330
QY      347 FNT-----FFRHGASPSYSSAVWYLTOLRRGMINEFKPNWYLDPTAKNYVRPIYLA 401
Db      331 -NTGIGQMFWGDHASYPFHSHDSWFTITENIKWG---KAPDIDYALVAKYNREDIMPA 386
QY      402 AKELVAESKAKADEFPADTSIKPSQNFIDKVPEDANKPNDYLAFAI 449
Db      387 AKDL---GVA---DLPASTS--RGKETFPDGKV-FPEPNSAYLESISI 426

```

RESULT 3  
 H98248  
 nitrate-binding protein ntrA precursor, periplasmic [imported] - Agrobacterium tumefaciens  
 C|Species: Agrobacterium tumefaciens  
 C|Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
 C|Accession: H98248  
 R|Goodner, B.; Hinkle, G.; Gatlung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Molim, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Matrelz, B.,  
 Science 294, 2333-2328, 2001  
 A|Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A|Reference number: A97359; MWID:21608551; PMID:11743194  
 A|Accession: H98248  
 A|Status: preliminary  
 A|Molecule type: DNA  
 A|Residues: 1-471 <KUR>  
 A|Cross-references: UNIPROT:Q8U925; GB:AE007870; P|DN:AAK9514.1; P|D:g15159390; GSPDB:G  
 C|Genetics:  
 A|Gene: AGR\_L\_1886  
 A|Map position: linear chromosome

Best L

Query Match	25.7%;	Score 627.5;	DB 2;	Length 471;
Best Local Similarity	35.5%;	Pred. No. 3.1e-38;		
Matches 166; Conservative	70;	Mismatches 171;	Indels 61;	Gaps 19;

A;Map position: linear chromosome

QY 1 MMTTIRSS-SKKULL--TTSASLAWGLTAPD--VQAVKLEKEDLKRGSTIKLTDMAPL 55  
 Db 42 MKKISGIVSRRIKLTMTYTAALVAVRTAFSGAFAPATAPEPKAGALGFTALTDMAPL 101  
 QY 56 AVAAEKGFPEDEGL-FVQLEAQAQNMKVMDRVVNG---ELDGSMTLAPAPLAASVSGT 110  
 Db 102 IIAAEKGLFAKHGMPEVEVLKQKSGKATRNVLVLGASNGIDGHHILTPMYLMTGKVT 161  
 QY 111 KADIEVPFS---MGFNNAITVSNELIWHQMKPNIPLEGKRVPHIKADYLPKVVEXYKA 166  
 Db 162 QNNVPVPMITILARLINDSGOISVAKF---YAEFGVQDASK-----LKAPEKKKA 209  
 QY 167 EKKPFMMANTFPAGSINIKLRYWLAGGINPGYISPPQDIDSGQIGADALLSVTPPQWPS 226  
 Db 210 DKKELKAAMTFPGGTGTDHLIRYWLAAAGID----PDKDVS-----TIVPPQWVA 256  
 QY 227 TLEAGTIFGFCYCGEPPNQOAVFPGIGVPTVTEELMTDTEKVPFGVTKQMAEKYENTYLA 286  
 Db 251 NMKVGGMDFVCYGEPPNEQLVNOGIGFTACTTGELMKGHEPKALGMRADWVEKQINATKA 316  
 QY 287 VTKALIRAAIWLADADNNKRKEAIEMLAQKQYGVADVEVLAASNMGTFEYEKDRLALPD 346  
 Db 317 LLMAVMEAQW--CDEMAKKEEMSTILGRQWFNVPPKQDVLGRKGNINY--GNGRVLE- 371  
 QY 347 FNT-----FFRGASIPSYSSAWLYLTOLRRKMGINEFFKDPNMYLTDRKANYRPPIYLA 401  
 Db 372 -NTGLQMKWQOHAASYPFHSHDSWFTTENIRWG---KFADPTDVKALVAKVNRREDIMRAA 427  
 QY 402 AAEIWAEGKAKADEFPADTSIKPSQNPFTDKVPFDANKPNDYLAKFAL 449  
 Db 428 AKDL---GVA---DLPASTS-RGKETFFDQKV-FDPENPSAYLESLSI 467

RESULT 4  
 E95928  
 probable nitrate transport protein nrtA [imported] - *Sinorhizobium meliloti* (strain 1021)  
 C:Species: *Sinorhizobium meliloti*  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
 C:Accession: E95928  
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrester, J.; Chain, P.; Vorholter, F.J.; Herman  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo  
 A:Reference number: A95842; NCID:21396508; PMID:11481431  
 A:Accession: E95928  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-441 <KUP-  
 A:Cross-References: UNIPROT:Q92VK8; GB:AL591985; PTDN:CAC49093.1; PID:g15140578; GSPDB:G  
 A:Experimental source: strain 1021, megaplasmid pSymb

R.Galibert, F.Flinan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubber, P.; Chaïn, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L. Hyman, R.W.; Jones, T. Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, M.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K. Article: The complete genome of the legume symbiont *Sinorhizobium meliloti*. A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: nrch, SMB21114

A:Genome: plasmid

[illegible]

```

RESULT 5
ADJ037
Hypochemical protein nrtc [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: ADJ037
R:Wood, D.W.; Seethal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McEllell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: ADJ037
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-431 <KUR>
A:Cross-references: UNIPROT:O80924; GB:AE008689; PIDD:AA144714.1; PID:g1742345; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: nrtc
A:Map position: linear chromosome

```

```

Query March 151, Conservative 25.0%; Score 611.5; DB 2; Length 431;
Best Local Similarity 35.5%; Pred. No. 4.1e-37;
Matches 151; Conservative 69; Mismatches 166; Indels 39; Gaps 12;

Dy 34 VGKLEKEDLKEGFIKLTIDMAPLAAVAAEKGFPEDEGLFPQLEAQANWVMDRVVNGELD 93
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 24 VGSQRQKILRGRGFIPIVDASVLIAAERGFADREGLLIDLVKQVSMANVRRLAFRGFDI 83
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 94 SHMLAPAPLAASVSGPTKADIEV--PFSMGFNQAITYSNSIETHQMKENIP--EGGKPVHP 151
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 84 AHMLSPMPVASMGLGSGNSPSTPTTPFSIGRGNATITSTRFAEMKALTGSETAGALEN 143
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 152 IKADYLKRVYAEKVAEBK--PENNMTPTPAGSHNKLRYVTLAAGINPGYSPPODISQI 210
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 144 ARA--LKLVLIDDMRAREBAPPTLGMTYPPSSHNTAEFRWYLAAGSIIHPH----- 190
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 211 GADALLSTVPPQNPSTLEAGTIFGYCVGBEPMNOQAVFGKGVIVTDEELMKDTPKEVF 270
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 191 --DYKLVVPPPTMSDALLAAGATDGFVYGAIPWNI--VLAERGVGRVYAAKQDIMPBAPEVYI 248
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 271 GVTYKQMAEKYPTNYIATVKALIRAAIWL--ADNNKNRKEALIMLAQKQYVADVEVLAAS 329
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 249 GMRFEMAESQETVYGRLLTLALDAASWCDLADNHDALSGA---LADPRYICAPQSIIRRV 305
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 330 MNGTFEYE--KDDKCALPDENTFFRHGASYSYSSAVWYLTQLRRWGMINEFKPDNNWYIDT 388
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 306 LAGEFSIDSQGNRRVRIEKEYTFHGDHANYPROSOSLMTYSQMIRWG---QAEISETGVNA 362
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 389 AKNNYRPDIYLAAKELVABGKAKEADPPADTSIK-----PSQNFIDKVPDPADAKPNY 444
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 363 ALSYRPDIYRAA---LGDGKA-----PGADIRIRIEGQDEGDRFVDFVDPADIAGYV 413
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Dy 445 AKPAI 449
      ::|::|
Db 414 NSPAV 418

```

RESULT 6

G988248

hypothetical protein AGR\_L\_1881g1, [imported] - Agrobacterium tumefaciens (strain C58, C

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C:Accession: G988248

R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman

A: Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359, MUID:21608551, PMID:11743194

A:Accession: G988248

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-431 <KUR>

A:Cross-references: UNIPROT:Q8U924; GB:AE007870; PIDN:AAK69513.1; PID:G15159389; GSPDB:Q

C:Genetics:

A:Gene: AGR\_L\_1881g1

A:Map position: linear chromosome

Query Match 25.0%; Score 611.5; DB 2; Length 431;

Best Local Similarity 35.5%; Pred. No. 4.1e-37;

Matches 151; Conservative 69; Mismatches 166; Indels 39; Gaps 12;

Qy 34 VGKLEKEDDKETGFKYKLDMAPLVAAREKPFEDSGTVOLEAQNMYKVMRDVYNGELDG 93

Db 24 VGSRRQKILTRGRFLPVDASVLLAAAEFGFADRRGLTLIDKVDVSMANVRRLFLRPQDI 83

Qy 94 SHMLAPAPLAASVGFGTKADIEV-PSFMGFNGAIIYTSNEIMHQMKENIPL-EGSKDVHP 151

Db 84 AHMLSPMPVASMIGLGSNPSPITIRFSLGRGNAITLSTRFLPARKALTLGSEFAGALEN 143

Qy 152 IKADYLRPVVEKYAAEBG-PNNMAMTPPAGSHNKLKATYVLLAAGINQGYSSPPDISGQI 210

Db 144 ARA--LKIIVDDMKAREBAPPTIGMTVPFSSHNYEFPYVLAAGIHPDH----- 190

```

QY 211 GADALLVTPPPQMPSTLEAGTIFGVCSEPMNOQAVFKIGIVPITDELMKDTPEKYE 270
DB 191 --DYKLVVPPPTSDALAGALDGFVGAPMNIVAERVGRIVAAKODLMSABEKYI 248
QY 271 GVTQMAEKYPNTYLAATKALIRAAIWL-ADNNKRRKEAIEMLAQOYGVADVEVLAAS 329
DB 249 GMPPEMAESQOETVGRLLTLALDAASWCDLADNHDLSGA---LADPRYIGAPQSIRRY 305
QY 330 MNGTFERE-KDDGRALPDPTFFRHGASPSYSASAVVYLTQLRRKGINEFKDNNYLT 388
DB 306 LAGEFSDSQGNRRVIEKYTFPHGDHANYPRQSQSLMYSQMRWG---QAEISETGVNA 362
QY 389 AKVVPDYLAAKEIVAEKKAEDFPADTSIK---PSQNFIDKVPFDANKENDLY 444
DB 363 ALSAVPDIYRAA-----LGDGKA-----PQDADIRIEGDEGRFVDFGVPFDADLAGY 413
QY 445 AKFAI 449
DB 414 NSFAV 418

```

## RESULT 7

```

AG1882
nitrate transport nitrate-binding protein ntrA [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: AG1882
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yaeuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1882
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-440 <KUR>
A:Cross-references: UNIPROT:Q44292; GB:BA000019; PIDN:BA072566.1; PID:g17129954; GSPDB:G
A:Experimental source: strain PCC 7120
C/Genetics:
A:Gene: ntrA

```

```

Query Match          25.0%; Score 610.5; DB 2; Length 440;
Best Local Similarity 33.4%; Pred. No. 5e-37;
Matches 188; Conservative 71; Mismatches 173; Indels 71; Gaps 17;

QY 8 SSKLLLT---LSASLAWGLTI-----ADPGAAG-KLEKEDLKF 44
DB 5 SRRKFLFTTGAANAASILVHGCTSNQSASATTGEOAPSAAPANVSAANAAPKVEETKAL 64
QY 45 GFKLTDMPPLAVAAEKGFEDGLF-VQLEAQANKVWDRY---VNGELDSHMLAP 99
DB 65 GFPLDADAPLTAKEKGFPAKYGMTDIEVTKQKSPVTRDKNLKISSGGIDGAHITLSP 124
QY 100 APLAASVGFSTKADIEVPSMGFNAGATIVSNEIWHQMKNPILLEGKRPVPIKADYLP 159
DB 125 MPTLMTI--NDKVPMTILARLNTNGQALISAEKF---KELNVLE-----SKSLND 170
QY 160 VVEKYRAEKGFPFNAMTFPAGSHNIKLRVYLAAGINPGYSPPODISQIGADLLAST 219
DB 171 AATKAKADKALMGITPFGTHTLWRYVWLAAGINP-----DQDVLEAV 217
QY 220 PPPQMPSTLEAGTIFGVCSEPMNOQAVFKIGIVPITDELMKDTPEKVFQVTKQMAEK 279
DB 218 PPPQMPANMKVNTVDGCGEPPNQAQVNOKIGYSAIVGELMKDPEKAFSRKQWMIQ 277
QY 280 YPNTYLAATKALIRAAIWLADNNKRRKEAIEMLAQOYGVADVEVLAASMNGTFEYED 339
DB 278 NPNAQAQALIMAILAQQM--CDKAEKKEWCKICDRKYPFNVAADAIIRAKGNIDY--G 333
QY 340 DKRALPDF--NTFFRHGASPSYSASAVVYLTQLRRKGINEFKPDNNYLTDAKNYRPD 396

```

```

DB 334 DGRKEQNFARHMKFWADNASYPYKSHDIWFLTEDIRMGYLPK---DTKQDILVNGVNKED 390
QY 397 IYLAAKEIVAEKKAEDFPADTSIKPSQNFIDKVPFDANKENDLYAKFAI 449
DB 391 LMKKAQ--ALGVADAE-IPASS-RGVEYTF-DGVAFDPEKPEEYLSLKI 437

```

## RESULT 8

```

AG1882
nitrate transport ATP-binding protein ntrC [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: AG1882
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yaeuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG1882
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-657 <KUR>
A:Cross-references: UNIPROT:Q8Y276; GB:BA000019; PIDN:BA072568.1; PID:g17129956; GSPDB:G
A:Experimental source: strain PCC 7120
C/Genetics:
A:Gene: ntrC
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

```

```

Query Match          24.0%; Score 587; DB 2; Length 657;
Best Local Similarity 35.8%; Pred. No. 4.6e-35;
Matches 150; Conservative 74; Mismatches 139; Indels 56; Gaps 17;

```

```

QY 39 KEDLKGPIKLTDMPPLAVAAEKGFEDGLF-VQLEAQANKVWDRYVNGELDSHML 97
DB 276 KAVIEIGFMPPLTDSAPLVAKEKGFPAKYGLDNVILNANMWAQATATGVTGKLLAAQMV 335
QY 98 AAPLAASVGFSTKADIEV--PFSMGFNAGATIVSNEIWHQMKNPILLEGKRPVPIKAD 155
DB 336 AGMPPLALTUGASQPTPTVINALNLSRANATTFEKRILYNGVRSL-----AD 383
QY 156 YLKPVEKYKAEKGFPMNMTFPAGSHNIKLRVYLAAGINPGYSPPODISQIGADAL 215
DB 384 LKAVID--SSPDQLTLTGIVHSAQMNLILRYWLAAGID-----PRDVS----- 427
QY 216 LSVTPPPQMPSTLEAGTIFGVCSEPMNOQAVFKIGIVPITDELMKDTPEKVFQVTKQ 275
DB 428 LTVIPPTQWVSQLKAGNIDGYCAGEPMNYQAVHDDLGVAAATALIWGQPKKVLGVRED 487
QY 276 WAKEYPNTYLAATKALIRAAIWLADNNKRRKEAIEMLAQOYGVADVEVLAASMNGTF- 334
DB 488 WAQKPEPTLYNLVYALIAECK--CDDLARNEBELICREPY-----LDVNPAYRSGET 541
QY 335 -EYKED--KRALPDPTFFRHGASPSYSASAVVYLTQLRRKGINEFKPDNNYLTDAK 390
DB 542 DEYDGDGTPPOLTAANYQFYLNTKNTYDNRTEILMITQMRWG-LTFP-PGNW-VEITE 598
QY 391 NVYRPDIYLAAKEIVAEKKAEDFPADTSIKPSQNFIDKVPFDANKENDLYAKFAI 449
DB 599 RVCRTDITFGAARDL-----GLLDIGEDPI---HLFDGL-FNPSPEIYLSLEI 646

```

## RESULT 9

```

C95300
ntrA-type periplasmic nitrate transport binding protein, probable Sma0585 [imported] - S
C/Species: Sinorhizobium meliloti
C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C/Accession: C95300
R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barlow-Hubler, F.; Bows
., Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432

```

A:Accession: C95300  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-430 <KUR>  
A:Cross-references: UNIPROT:O92221; GB:AE006469; PIDD:AAK64965.1; PID:G14523390; GSPDB:G  
A:Experimental source: strain 1021, megaplasmid pSYMA  
A:Galibert, F.; Ffinan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chail, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kise, B.; Komp, C.; Lelaure  
beault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A:Reference number: A96059; MUID:21366234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: SMA0585  
A:Genome: plasmid

Best Local Similarity 32.7%; Pred. No. 1,4e-32;  
Matches 148; Conservative 68; Mismatches 179; Indels 57; Gaps 16;

QY 6 RSSSKLLTLTSLAVM -GLTAPDVGAAGKLEKEDEKGFITLDTMAPLVAAGKGF 64  
DB 16 RSLAMGAAATLAAKAKAFPGAAHAGAP---EVAKARIGFALTDDSSPLIAKRGF 71  
QY 65 EDEGL-FVQLEAQAQNMKVMDRVV---NGELDGHMLAPAPLAASVGF---GTKADIEV 116  
DB 72 AKGMPDIEVVKQASMAATRDNLVLGAERGIIDGAILTPTLMTTGTITGGAFFPMYI 131  
QY 117 PSMGFNGNAITVSNELIWHQMKNPILLEGKPVHPKADYLKPVKEKYAKGKPFMMAMT 176  
DB 132 LALNTNGQGISVGNNDL-KSVKVLNSAGA-----KAKFL-----QMKAGNIAYKAMT 179  
QY 177 FPAASHNIKLRVLAAGGIRPGYVSPPODISQIGADALLSTPPQMPSTLEAGTIFG 236  
DB 180 FPGTIDHLMTRVWLAAGGIRP-----DVDVSTIVLPQPMANMKAQTQDAF 226  
QY 237 CVGEPMNQAVFGIGVPIITDEELMKDTEPEKVFQVTKQAEKYPNTYLAATKALIRAAI 296  
DB 227 CVGEPMNQAVLVNKGVTACLTISELMNHPEKALGMRASWVDKYPRAAQALTAAGVDAQM 286  
QY 297 WLDADNNKREKALEMLAQOYVGADVEVLAASMGTFEYKDKRAL---PDPTFFPR 353  
DB 287 W--CDKRAANPQMCISVGRQYVNVPMGDILPRLQGTVDY--GDGRTLKNSPRMKFMAD 342  
QY 354 GASVPSYSAVMWYLTQLRKGMINEPEKPDWYLDTKANVRRPITYLAAAEVLAEGKAKA 413  
DB 343 NASPFKSHDLWFLTEIDIRWGLVPQKTNTKALVD--KVRNSDIWRPAAKSIQSG--- 395  
QY 414 EDPADTSIRKSNFPIDKVPDPANKRNDYLA 445  
DB 396 ---PAGDS-RGVERFEDGKV-FDPANFGAYLA 422

RESULT 12  
S58738  
nitrate-binding protein nrtA precursor, periplasmic [similarity] - Phormidium lamiosum  
C/Species: Phormidium lamiosum  
C/Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 09-Jul-2004  
C/Accession: S58738; S56641; S62124  
R/Merchan, F.; Kindle, K.L.; Liama, M.J.; Serra, J.L.; Fernandez, E.  
Plant Mol. Biol. 28, 759-766, 1995  
A/Title: Cloning and sequencing of the nitrate transport system from the thermophilic, *Phormidium* sp. PCC 7942  
A/Reference number: S58738, MUID:95375238; PMID:7647306  
A/Accession: S58738  
A/Molecule type: DNA  
A/Residues: 1-442 <MER>  
A/Cross-references: UNIPROT:Q51880; EMBL:Z19598; NID:91154890; PID:CA879656.1; PID:e218  
R/Merchan, F.; Pileto, R.; Kindle, K.L.; Liama, M.J.; Serra, J.L.; Fernandez, E.  
Plant Mol. Biol. 27, 1037-1042, 1995  
A/Title: Isolation, sequence and expression in *Escherichia coli* of the nitrite reductase  
A/Reference number: S56640; MUID:95284340; PMID:7766873  
A/Accession: S56641  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-67, 'A', 69-442 <MEM>  
A/Cross-references: EMBL:Z19598  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C/Genetics:  
A/Gene: nrtA  
C/Keywords: nitrate transport; periplasmic space

Query Match 22.3%; Score 545.5; DB 2; Length 442;  
Best Local Similarity 33.3%; Pred. No. 2.9e-32;  
Matches 145; Conservative 67; Mismatches 175; Indels 49; Gaps 14;

QY 27 IAP-DVGAAGKLEKEDLKRGFILTDMAPLVAAGKGFEDGL-FVQLEAQAQNMKVMD 84  
DB 44 LAPAPVAAADTPEVTAKGLICPDGLAPLIIAKRGFLKAPKGMPEVLEVKQASMAVTD 103

QY 85 RVV---NGELDGHMLAPAPLAASVGFETKADIEVPS---MGFNGNAITVSNELIWHQ 136  
DB 104 NIELSGGGGIDGAILITPMPYLSAGTITKKNQVPPNNIILARLVNNGGICLANTY--- 160  
QY 137 MKPNIPLEGKRPVHPKADYLKPVVEKYAKGKPFMMAMTPPAGSHNIKLRVWLAAGGIN 196  
DB 161 ---KDLKIGTSSDLKEAFAKA-----KAEKEIKAAVTFFGCTHLMRWYLSAGGID 211  
QY 197 PGYSPPODISQIGADALLSTPPQMPSTLEAGTIFGCVGEPMNQAVFKGIGVPI 256  
DB 212 ---PDKDIS-----TIVPPQVAVANKVMNMEFECVGEPMAPQTAVQGLGNM 258  
QY 257 TDEELMKDTEPEKVFQVTKQAEKYPNTYLAATKALIRAIWLDADNNKREKALEMLAQK 316  
DB 259 TTGELMKHPEKAFAMRADWVEGHHKAKALLMAVQEAQIW--CDDPANKEMCGVSGR 316  
QY 317 QYVGADVEVLAASMGTFEYKDKRALPDPTFFRHGASYSYSAAWYLTQLRKGM 376  
DB 317 EWFKPVVEDILERSKGNFDLGVRLQENSPILMKFMNDASYPFKSHDLFLTEIDIRWGL 376  
QY 377 NEFKDNNWYLDPAKNVRRPDIYLAAKELVAEGKAKAEDFPADTSIKSONFPIDKVP 436  
DB 377 PADDTKALVDA---VNREDLMREAKAIGQEA-----IPASTS-RGVETFP-DGVKPD 426  
QY 437 ANKPNDYLAFAIGLK 452  
DB 427 PENPSAYLS--ALKKK 440

RESULT 13  
AH2165  
bicaarbonate transport ATP-binding protein cmcC [imported] - Nostoc sp. (strain PCC 7120)  
C/Species: Nostoc sp. PCC 7120  
A/Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C/Accession: AH2165  
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saesomco, S.; Matanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena* sp. PCC 7120  
A/Reference number: AB1807; MUID:21595285; PMID:11759840  
A/Accession: AH2165  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-667 <KUR>  
A/Cross-references: UNIPROT:Q8Y147; GB:BA000019; PID:BA874578.1; PID:g17131973; GSPDB:G  
A/Experimental source: strain PCC 7120  
C/Genetics:  
A/Gene: cmcC  
C/Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 22.0%; Score 538.5; DB 2; Length 667;  
Best Local Similarity 33.3%; Pred. No. 1.7e-31;  
Matches 140; Conservative 70; Mismatches 158; Indels 53; Gaps 14;

QY 30 DVGAAGKLEKEDLKRGFILTDMAPLVAAGKGFEDGL-FVQLEAQAQNMKVMDRVV 88  
DB 270 DVARRG-LEKNVLIGFLPRLTACAPLAAGKGFETGGLDEVLVRETSWRGIDGKMG 328  
QY 89 GELDGHMLAPAPLAASVGFETKADIEV--PSMGFNGNAITVSNELIWHQKNPILLEGK 146  
DB 329 GYIDAQAQMPGPMWLTGSHDNGPLPVYALTITRNAGNATILARFYD-----EGV 380  
QY 147 KPVHPIKADYLKPVVEKYAKGKPFMMAMTPPAGSHNIKLRVWLAAGGINPGYSPPODI 206  
DB 381 RSLSDPK-NYL-----LRTDQRHIMGVVHNSMNILLRWYLAAGGIDP----- 424  
QY 207 SQGIDADALLSTPPQMPSTLEAGTIFGCVGEPMNQAVFGIGVPIITDEELMKDTP 266  
DB 425 ---DLDVDKRTIPRAQVADLQNSIDGICVGEPMNRAAVENGTITADLEWVLGHP 480  
QY 267 EKVFGVTKQMAEKYDNTYLAATKALIRAIWLDADNNKREKALEMLAQOYVGADVEVL 326



Db 481 GKVLGVREDNARERYPNTHIALTKALLEACEY--CSRPNENVEARIVAGRDVSTDDLYI 538  
Qy 327 AASNGAFPEYEKDKRALPDP--NTFRRHGA-STPSSASAVWYLTQLRRMGINEFKPDN 383  
Db 539 QLEDPNSLVCDID--HPLRDVAHQFPAESALNRPSTEQIWMISQLARWG--DTPPRN 594  
Qy 384 WYLDTAQNVYRPDIYLAAXELVAEGKAKEDFPADTSIKRSQNFIDKVPEDANKENDY 443  
Db 595 W-VEVERVCRKRVSTFAREL-----GLDLSYTRQPIELFDGTPFNADDPYAY 642  
Qy 444 L 444  
Db 643 L 643

RESULT 14  
S75959  
nitrate transport protein C-1 - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: protein glr0043  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 16-Aug-2004  
C:Accession: S75959  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S75959  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-667 <KAN>  
A:Cross-references: UNIPROT:Q55462; EMBL:D64006; GB:AB001339; NID:G1001291; PIDN:BA01080  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Gene: ntrC-1  
C:Superfamily: ATP-binding cassette homology  
C:Keywords: ATP; nucleotide binding; P-loop  
F:25-215/Domain: ATP-binding cassette homology <ABC>  
F:42-49/Region: nucleotide-binding motif A (P-loop)

Query Match 21.9%; Score 534.5; DB 2; Length 667;  
Best Local Similarity 31.7%; Pred. No. 3.3e-31;  
Matches 146; Conservative 73; Mismatches 154; Indels 87; Gaps 16;

Qy 10 KLLLTLSASIAVWGLTIAPDVGAAGKLEKEDLKFGFIKLTDMAPLVAAEKGFEEDEGL 69  
Db 261 KKLRAKKTITAIARHG-----LEKYNLELGIYVPLVACPLVAOEGKFFFAAGHL 308  
Qy 70 -FVQLEAQAAMKVMRVNNGELDGHMLAPAPLAASVGFGRKADIEV--PFSMGFNGNA 126  
Db 309 DEVLAVRETSWIRGIVDGIAGYLDGAQMPAGMPTWLAAGVREGSIVVSAALMTTRGNA 368  
Qy 127 ITYSNEMWQMKNPILLEGKPNHPIADYLKPVVEKYKABGKPFNNMFTFPAGSNIKL 186  
Db 369 ITLSKLLYD-----GIYTAE-DEROLL--ASDGRHTLTGMVHPSSMNLIL 413  
Qy 187 RYMLAAGINPGYSPPODISGQIGADALLSVTPPOMPSTLEAGTIFGYCVGEPPNQQA 246  
Db 414 RYMLAANNIP-----DRVHLKTIIPPAQVADLKAGITIDGCVSSEPMURA 460  
Qy 247 VFKGIVGVITDEELWKDTEPEKVGVTQKQAEKYPNTYLAATKALRAATLADANNKNR 306  
Db 461 SMEGAGSFIATDLEIWMNHPGKVLGVREDWALAHPTNTHVALYKALLBAAYC-ADPN-HE 518  
Qy 307 KEAIEMLAKQOYVADDEVL-----AASNGTFEYEKDKRALPDNTFRRHGAAYP 358  
Db 519 MEIRELLATROYLSTNIDYIHLGDPGRCRLGNPVEY-----SHHLFGDQFNRP 569  
Qy 359 SYSASAVVYLTQLRRMGINEFKPDNWTLDTAQNVYRPDIYLAAXEL-----V 406  
Db 570 SRTEHLMMTQMARKGDI-P--FRANW-VEILERVCRGVSTARELGYVNTYORPIAL 626

Qy 407 AEGKAKAEDFPADTSIK-----PSQNFIDKVPEDANKP 440  
Db 627 FDGKV-----FNADDPYALYNQTVIHRNFTIAEVLNPTP 662

RESULT 15  
S77389  
nitrate transport protein A-2 - *Synechocystis* sp. (strain PCC 6803)  
N:Alternate names: nitrate transport 45K protein; protein glr1450  
C:Species: *Synechocystis* sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: S77389  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp.  
A:Reference number: S74322; MUID:97061201; PMID:8905231  
A:Accession: S77389  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-446 <KAN>  
A:Cross-references: UNIPROT:P73452; EMBL:D90906; GB:AB001339; NID:G1652492; PIDN:BA01174  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Gene: ntrA-2

Query Match 21.5%; Score 525; DB 2; Length 446;  
Best Local Similarity 28.5%; Pred. No. 9.3e-31;  
Matches 139; Conservative 78; Mismatches 186; Indels 84; Gaps 15;

Qy 1 MKTIIRSSKKLLTASIAVWGL-----TIAPVGAAGKL 37  
Db 1 MSNFSRSTRKRMFTAGAA-AIGGVVILHCTSTTTSTGTGSSSTDQALSPVBENAP 59  
Qy 38 EKEDLKFGFIKLTDMAPLVAAEKGFEEDEGL-FVQLEAQAAMKVMRVN-----NGELD 92  
Db 60 EVTTAKLGFIALTDAAPLIIAKEKGFYAKGMDEVELKQASWGTTRDNLVLGASAGSID 119  
Qy 93 GSHMLAPAPLAASVGF-----GTRKADIEVPFSMGNGNAIYVSNEMWQMKNPILLEGKPY 149  
Db 120 GAHILTPMPLITMGVTGDKPTMYILARLANNQGIOLGNVY-----KDLKVGTDPA 172  
Qy 150 HPIKADYLKPVVEKYAEGKPFNNMFTFPAGSNIKLRYMLAAGINPGYSPPODISGQ 209  
Db 173 APIKEAPAKVTDK-----VAMTFPGGTHDMIRYMLAAGNEP----- 211  
Qy 210 IGADALLSVTPPOMPSTLEAGTIFGYCVGEPPNQQAVERKGIQVPIITDEELWKDTEPEKY 269  
Db 212 -GDSFTIVPPAQVAVNKKVAMBSFCVGEPPPLQTVAGVGYALLTGQLMKDHPKXA 270  
Qy 270 FGYTRQAEKYPNTYLAATKALRAATLADANNKRAIEMLAKQOYVADDEVYLAAS 329  
Db 271 FGRRADWVQNPQKAKALMAVMEAOQW--CQOAEKKEECQITISREWFKYPFEDIIDR 328  
Qy 330 MNGTFEYEK-----DDKRALPDNTFRRHGAASPSASAVWYLTQLRRMGINEFKPDW 384  
Db 329 SKGIYVFGNGOEFEEDEIEM--QKTVWDASTPYFSHQWFLTEIRNGYLPASTDTYKA 385  
Qy 385 YLDTAQNVYRPDIYLAAXELVAEGKAKAEDFPADTSIKRSQNF--FIDKVPEDANKPND 442  
Db 386 IVD---KNREDLMEBAQAL-----EVPAQDQPS-----SPSRGIEFTFDGTFDPENPQA 434  
Qy 443 YLAKEAI 449  
Db 435 YLDSLKI 441

Search completed: September 16, 2005, 01:09:58  
Job time : 46 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 00:49:17 ; Search time 174 Seconds

(without alignments)  
1365,544 Million cell updates/sec

Title: US-10-689-200-2

Perfect score: 2443  
Sequence: 1 MKTIISSSKKLLTUSASL.....AKPAIGLKQYVAGKRVVD 464

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : uniprot\_03:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1597	65.4	462	2	08KY55 azospirillum
2	1586	64.9	475	2	07MDT5
3	1572	64.3	487	2	08D6X8
4	1564	64.0	454	2	087HA7
5	1257.5	51.5	553	2	07UYV1
6	651	26.6	670	1	NRTC_SVNY3
7	627.5	25.7	430	2	08U925
8	627.5	25.7	471	2	07CTL1
9	616.5	25.2	441	2	092VK8
10	611.5	25.0	431	2	08U924
11	610.5	25.0	440	1	NRTA_ANASP
12	610	25.0	429	2	07NEC4
13	594	24.3	437	2	07NKB1
14	587	24.0	657	2	08Y276
15	585.5	24.0	420	2	08Y2F5
16	582.5	23.8	439	2	089RG9
17	578.5	23.7	412	2	098H14
18	578	23.6	420	2	06D2V0
19	577	23.6	439	2	08DJ78
20	576	23.6	430	2	092Z21
21	575.5	23.6	418	2	048466
22	572	23.4	443	1	NRTA_SVNP7
23	567	23.2	654	2	08DJ76
24	566	23.2	425	2	098H12
25	559.5	22.9	388	2	089LH2
26	549.5	22.5	430	2	09AAI8
27	548.5	22.5	448	2	07N1U6
28	548.5	22.4	428	2	08DHF7
29	548	22.4	428	2	06WRT2
30	545.5	22.3	442	2	051880
31	542.5	22.2	666	2	07N1U3

32	540	22.1	655	2	07NKA9	07nka9 gloeobacter
33	538.5	22.0	667	2	08Y747	08y747 anabaena sp
34	536.5	22.0	663	2	055107	055107 synchococc
35	534.5	21.9	667	2	055462	055462 synchococc
36	531	21.7	404	2	088L41	088L41 pseudomonas
37	525.5	21.5	462	2	06N722	06n722 rhodopsendo
38	525	21.5	403	2	0883P3	0883P3 pseudomonas
39	525	21.5	446	1	NRTA_SVNY3	P73452 synchocyst
40	521.5	21.3	533	2	06W1F8	06w1f8 synchococc
41	510.5	20.9	459	2	0891A5	0891a5 bradyrhizob
42	509.5	20.9	559	2	07U3E5	07u3e5 synchococc
43	506	20.7	386	2	06N3I6	06n3i6 rhodopsendo
44	502.5	20.6	458	2	08Y749	08y749 anabaena sp
45	494	20.2	402	2	0912V6	0912v6 pseudomonas

## ALIGNMENTS

## RESULT 1

08KY55 PRELIMINARY; PRT; 462 AA.  
AC 08KY55;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Nitrate transporter substrate-binding protein.  
GN Name=naeF;  
OS Azospirillum brasilense.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
OC Rhodospirillaceae; Azospirillum.  
OX NCBI\_TaxId=192;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SP245;  
RA Steenhoudt O., Piacsek D., Verreth C., Vanderleyden J.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF283498; AAM73543.1; -  
DR InterPro; IPR002110; ANK.  
DR PRINTS; PR01415; ANKYRIN.  
SQ SEQUENCE 462 AA; 50270 MW; 25C86DF070421EBC CRC64;

Query Match	Score	Length	DB 2;	Length	462;
Best Local Similarity	65.4%;	Pred. No. 6.3e-112;			
Matches 301; Conservative	60;	Mismatches 92;	Indels	6;	Gaps 4;
7 SSSKKLLTILASIAVWGLTIAPDVGVKLEKEDLKFIFIKLTDMAPIAAAEKGFED 66					
6 ASPTLLLSAAATLA---LMLSSAQAAPLDVEKDKLGFILKTDMAPIAAAEKGFED 62					
67 EGIPLVQLAQAQWKKVMDRVNNGELDGSIMLAAPAPIAASVGFQTKADIEVPSMGFNGNA 126					
63 EGIPLVQLAQAQWKKVMDRVNNGELDGSIMLAAPAPIAASVGFQTKADIEVPSMGFNGNA 122					
127 ITVSNFIMHOMKPNIFL-BGKRVPIKADYIKPVVEKTKAKGKPPNMAATPPAGSHNTK 185					
123 ITLSNEVWERKPNLEKPGDKPLPIKADALKPVAIQRAAGKPPNMAATPPAGSHNTK 182					
186 LRYWLAAGINPGVYPPPODISGQIGADALSTVPPQMPSTLEAGTIFGVYCGEPMNQO 245					
183 LRYWLAAGINPGVYPPPODISGQIGADALSTVPPQMPSTLEAGTIFGVYCGEPMNQO 241					
246 AVFKGIGVPIITDEILMKDTPKRVGVTKQMAEKYPNTYLAATKALIRAAIMLDADNNK 305					
242 AVFKGIGVPIITDEILMKDTPKRVGVTKQMAEKYPNTYLAATKALIRAAIMLDADNNK 301					
306 KREAIEMLAQKQYGVADVLAASNGTFRYKDKRALPDNTFPRHASTPSYSAYW 365					
302 RLEAVKILAKSEVGVADAVIANSMGTFRYKDKRALPDNTFPRHASTPSYSAYW 361					
366 YLTQLRRQMINEFKPDNNYLDTAQNVVPPDIYLAALKELVNEGKAKADFP-ADTSIRP 424					
362 YLTQLRRQMINEFKPDNNYLDTAQNVVPPDIYLAALKELVNEGKAKADFP-ADTSIRP 421					

QY 425 SONFIDKVPDPANKPNDYLAKFAIGLKGKOTVAGGKV 463  
DB 422 LDNGFIDGLAYDGRKRPVEYLTLPILGLKGQAVQGGLV 460

## RESULT 2

07MDT5 PRELIMINARY; PRT; 475 AA.  
AC Q7MDT5;  
DT 01-MAR-2004 (TREMBLrel. 26, Created)  
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Nitrate transporter system, periplasmic component.  
GN OrderedlocusNames=VVA0951;  
OS *Vibrio vulnificus* (strain YJ016).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; *Vibrio*.  
OX NCBI\_TaxID=196600;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14656965; DOI=10.1101/gr.1295503;  
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,  
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,  
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;  
RT "Comparative genome analysis of *Vibrio vulnificus*, a marine  
pathogen."  
RL Genome Res. 13:2577-2587(2003).  
DR EMBL; AF005348; BAC6977.1; -.  
KW Complete proteome.  
SQ SEQUENCE 475 AA; 53801 MW; B61BED1915DC367 CRC64;

Query Match 64.9%; Score 1586; DB 2; Length 475;  
Best Local Similarity 65.6%; Pred. No. 4.4e-11;

Matches 252; Conservative 60; Mismatches 85; Indels 8; Gaps 5;

QY 20 LAWGLTLPADV--GAVGKLEKEDLKFGFIKLTDMAPLVAAEKGFEEDEGLFVQLEAQA 77  
DB 32 LTSW-LAFAPPVLAEGAEKEDLKFGFIKLTDMAPLVAAYKGFEEDEGLVYTLLEAQA 90  
QY 78 NMKVNDRVNGLDGSNMLAPAPLAASVGFQKADIEVPFSMGFNQNAITVSNEIWHQM 137  
DB 91 NMVLLDRVIDGELDGAHMLAGOPLGATIGITQAVITAFSDMLNGNAITVSDVWQOM 150  
QY 138 KPNIPLEG-GKPVHPKADYLKPVVEKYKAEKGFPMAMTFPGSHNIKRLVWLAAGIN 196  
DB 151 KPMIAKOSDGKPVHPKADALKEVVTSYRQDGAFFMGWFFPVSTNYELRTWLAAGIH 210  
QY 197 PGYSPPO-DISQIGADALLSVTPPOMPSTLEAGTIFGVCGEPMNOQAVFKGIGVPI 255  
DB 211 PGYYAHKGDNSQINADVLSVTPPOMPATMEAGTIKGYCVGEPMNOQAVFKGIGVPI 270  
QY 256 ITDEELMKOTPEKVFQVTKQMAEKYPTNYLAVTKALIRAIWLDADNNKREAIEMLAQ 315  
DB 271 VTDYEIMKNNPEKVFQVADQMAEKYPTNTHIRVVKALIRAHWLDENDNNRRAEAVKLSRS 330  
QY 316 KOYVGADVEYLAASNGTTEYEKDDKRALPDENTFFRHGASYSYSAYWYLTQLRRQM 375  
DB 331 SEYVGADAEVYLAASNGTTEYEKDDKQVDFNVFPRYNATYYSDAIWTQLRRMQQ 390  
QY 376 INEFKPDNMYLDTAKNVYRPDIYLAAXELVAEGKAKAEDF--ADTSIKPSQNFIDK 432  
DB 391 IEHQKDSWYMDIAKQVYRPDIYQRAAEALIEEGTISASDFPFAKSGRRPQTHFIDQ 450  
QY 433 VEPDANKPNDYLAKFAIGLKGKQTV 457  
DB 451 IHYDGRSPNAVYLQQFAIGLKGSESL 475

## RESULT 3

08D6X8 PRELIMINARY; PRT; 487 AA.  
ID 08D6X8  
AC 08D6X8;

DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE ABC-type nitrate/sulfonate/dicarbonate transport system, periplasmic component.  
GN OrderedlocusNames=VW20393;  
OS *Vibrio vulnificus*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; *Vibrio*.  
OX NCBI\_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMC6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.;  
RT "Complete genome sequence of *Vibrio vulnificus* CMC6."  
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016809; AA007349.1; -.  
KW Complete proteome.  
SQ SEQUENCE 487 AA; 54515 MW; FE5D5F1C5B865F4 CRC64;

Query Match 64.3%; Score 1572; DB 2; Length 487;  
Best Local Similarity 65.1%; Pred. No. 5.2e-110;

Matches 289; Conservative 63; Mismatches 86; Indels 6; Gaps 4;

QY 20 LAWGLTLPADVGA-VGKLEKEDLKFGFIKLTDMAPLVAAEKGFEEDEGLFVQLEAQA 78  
DB 44 LTSWLAFAFSPVLAEGAEKEDLKFGFIKLTDMAPLVAAYKGFEEDEGLVYTLLEAQA 103  
QY 79 NMKVNDRVNGLDGSNMLAPAPLAASVGFQKADIEVPFSMGFNQNAITVSNEIWHQM 138  
DB 104 WKVLLDRVIDGELDGAHMLAGOPLGATIGITQAVITAFSDMLNGNAITVSDVWQOM 163  
QY 139 PNIPLEG-GKPVHPKADYLKPVVEKYKAEKGFPMAMTFPGSHNIKRLVWLAAGINP 197  
DB 164 PNIVKDSGKPVHPKADALKEVVTSYRQDGAFFMGWFFPVSTNYELRTWLAAGIHP 223  
QY 198 GYSPPO-DISQIGADALLSVTPPOMPSTLEAGTIFGVCGEPMNOQAVFKGIGVPI 256  
DB 224 GFYAHKGDNSQINADVLSVTPPOMPATMEAGTIKGYCVGEPMNOQAVFKGIGVPI 283  
QY 257 TDEELMKOTPEKVFQVTKQMAEKYPTNYLAVTKALIRAIWLDADNNKREAIEMLAQ 316  
DB 284 TDEYIMKNNPEKVFQVADQMAEKYPTNTHIRVVKALIRAHWLDENDNNRRAEAVKLSRS 343  
QY 317 QYVGADVEYLAASNGTTEYEKDDKRALPDENTFFRHGASYSYSAYWYLTQLRRQM 376  
DB 344 EYVGADAEVYLAASNGTTEYEKDDKQVDFNVFPRYNATYYSDAIWTQLRRMQQ 403  
QY 377 NEFKPDNMYLDTAKNVYRPDIYLAAXELVAEGKAKAEDF--ADTSIKPSQNFIDK 433  
DB 404 EHQKDSWYMDIAKQVYRPDIYQRAAEALIEEGTISASDFPFAKSGRRPQTHFIDQ 463  
QY 434 PPDANKPNDYLAKFAIGLKGKQTV 457  
DB 464 RYDGRSPNAVYLQQFAIGLKGSESL 487

## RESULT 4

087HA7 PRELIMINARY; PRT; 454 AA.  
ID 087HA7  
AC 087HA7;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Putative nitrate transport protein.  
GN OrderedlocusNames=VPA1058;  
OS *Vibrio parahaemolyticus*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; *Vibrio*.  
OX NCBI\_TaxID=670;  
RN [1]  
RP SEQUENCE FROM N.A.

RC	STRAIN-RIND_2210633 / Serotype O3:K6;
RX	MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA	Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA	Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA	Yasunaga T., Honda T., Shingawa H., Hattori M., Iida T.;
RT	"Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT	distinct from that of V. cholerae.";
RL	Lancet 361:743-749(2003).
DR	EMBL; AP005087; BAC62401.1; -.
KM	Complete proteome.
SQ	SEQUENCE 454 AA; 50650 MW; CCI741FBAIE2D751 CXC64;
Query Match	64.0%; Score 1564; DB 2; Length 454;
Best Local Similarity	65.0%; Pred. No. 1.9e-109;
Matches 293; Conservative 62; Mismatches 88; Indels 5; Gaps 5;	
Oy	13 LTLTSA-SLAWGLTIAPDVAGVGLLEKEDLKFGFKITKTDMAPLAVAEKGFPEDEGLFV 71 :::     :::     :::     :::     :::     :::     :::     :::     :::
Dd	6 MLKYSALSISVF--TAVASAKAQQLGPEELIEDLKFGFKITKTDMAPLAVAEKEGFPEDEGLFV 63 :::     :::     :::     :::     :::     :::     :::     :::
Oy	72 QLEAQAANKVVMVDVRVNGELDGSHMLADAPILAASVGFTKADIIEVPFSMGNGNAITVSN 131 :::     :::     :::     :::     :::     :::     :::     :::
Dd	64 TLEQAAMKVLLDRITIDELDGAAHMLAQPLIGATIGVTAKAEVLTARSMDLGNALITYSN 123 :::     :::     :::     :::     :::     :::     :::     :::
Oy	132 EIMWOMKENIPLE-GGKVPHPIKADYLKYVEYKYEKGKPPNNMATPBPAGSHNIKLRMYL 190 :::     :::     :::     :::     :::     :::     :::     :::
Dd	124 DTWEQMCKHIPIKQPGKRPVHPHKADSLKPVDVSRYDKSGKSPNMGVFPVSSHNYELRWL 183 :::     :::     :::     :::     :::     :::     :::     :::
Oy	191 AAGGINPGYBPP-Q-DISGOIGADALLSVTPPPQMPSTLENGTIFGYCVGEPPNQAVFK 249 :::     :::     :::     :::     :::     :::     :::     :::
Dd	184 AAGGIHPGFYAPDSGNSGOLDADVLTSVTPPPQPMPATMEGGTIKGYCVGEPPNQAVFK 243 :::     :::     :::     :::     :::     :::     :::     :::
Oy	250 GIGVIVITDEELMKOTPEKVGTVTKQMAKKPNYLIATVKALLIAIWLDDANNKNKREA 309 :::     :::     :::     :::     :::     :::     :::     :::
Dd	244 GIGTFVVDTVEIKSNPKPVGVSSKAAMEKPNHTIRVKKALIPAAHLIDENSPANRQEA 303 :::     :::     :::     :::     :::     :::     :::     :::
Oy	310 IEMLAQOKYGADVLEVLAASNMGTFEYEEKDKRALPDENTFRHGASYPSYSSAWYLTQ 369 :::     :::     :::     :::     :::     :::     :::     :::
Dd	304 VKMLAKSEYVGADADVIANSMGTGFEEYEGSKXPDPDPNVFFRNATYPYSDAILWTIQ 363 :::     :::     :::     :::     :::     :::     :::     :::
Oy	370 LRPMGINEFEPDNMYLDTAKNVRPDIYILAAVELVAEGAKAKADEPF--ADTSIKPSQ 426 :::     :::     :::     :::     :::     :::     :::     :::
Dd	364 MRNRGQISSEKSDMDVMVAKEVVRPDYIOOAAQHLIEDGLYSKDFPDFSABEDGRFPQ 423 :::     :::     :::     :::     :::     :::     :::     :::
Oy	427 NFPIDKVPFDANKPKVDYLAKFAIGKKGQTV 457    :::     :::     :::     :::     :::     :::     :::
Dd	424 THFIDNIYDGRREPNTYLEKFSIGLGKDXY 454    :::     :::     :::     :::     :::     :::     :::
RESULT 5	
Q7UYV1	
ID	PRELIMINARY; PRT; 553 AA.
AC	Q7UYV1;
DT	01-OCT-2003 (TREMBLrel. 25, Created)
DT	01-MAR-2003 (TREMBLrel. 25, Last sequence update)
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE	Nitrate transport ATP-binding protein.
CN	Name=nrcC; OrderedLocustName=NB370;
OS	Rhodospirillum rubrum.
OC	Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC	Planctomycetaceae; Firellula.
OX	NCHI_TaxID=117;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=N1;
RX	MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;
RA	Gloeckner F.O., Kobe M., Bauer M., Teeling H., Lombardot T.,
RA	Ludwig W., Gade D., Beck A., Borzum K., Heilmann K., Rabus R.,
RA	Schleutner H., Amann R., Reinhardt R.;
RT	"Complete genome sequence of the marine planctomycete Firellula sp.
RT	strain 1."
RL	Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR	EMBL; BX294133; CAD71540.1; -.

DR GO; GO:0005524; F:ATP binding; IEA.  
SQ ATP-binding; Complete proteome.  
KW SEQUENCE 553 AA; 60761 MW; 743DFAC3267B5D63 CRC64;  
  
Query Match 51.5%; Score 1257.5; DB 2; Length 553;  
Beet Local Similarity 50.7%; Pred. No. 3.2e-86;  
Matches 249; Conservative 74; Mismatches 133; Indels 35; Gaps 8;

QY KTIIRSSSKLLTLTSLAAVW-----GLTIAPDVGVGK----- 36  
DB 2 KAIIRTTKMLLGCVALAAAFPGCADSGVLEDELAALKVDISKIEVDNVPESAAVL 121  
QY 37 -LEKEDLKFGFIKLTDNAFLVAAEKGFFEDBEGLFVOLEAQANMKVMYDRVNGELDGSH 95  
DB 122 DLEKSDLTFFGFITKLTDCAPLVIAEKGYFDDEGLNTLVLETQSMKILLDVINGQLDGAH 181  
QY 96 MLAPAPLAASVGFTGKADIEVPSMGKNCAITYSNEIWHOMKNIP-LBGGRVHPYKA 154  
DB 182 MLAQPFGIGATIGVGTSPVIYAYSILDVNGGITNSNEVMAQMOWENDELKSPFKHPISA 241  
QY 155 DYLPKRVKEKYAE-GKPPNNAMTFPPAGSHNIKLRVTYLAAAGINGVYSPPODISGQIGAD 213  
DB 242 ASLKPRIYEDYIQDDPEPPGMVFVSITHNEIRYWLAAISGHGMTT-ESDIKIGFTDAQ 300  
QY 214 ALLSVTPPEPMPTSTLEAGTIFGVCVGEPMNOQAFAFKSIGVPIIDEEIAMKDPEKVGVT 273  
DB 301 VKLSITVPPQMPQRPQLBADIYMGVCVGEPMNQKAVVTGICPVTTNYDIWKNNPEKVGVT 360  
QY 274 KOMAKFPNTYLAAYTKALIRAAIWLD--DNKK--NRKALEMIAOKQYGADVIVLAAS 329  
DB 361 GEMAEKRPQTHTLAIVIKAILIRAGKMLDATTDSGKIWNREEAWEIISRDOYGAEEKVIGNS 420  
QY 330 MNCGFEVEKCDKRALPDNPFPFRHGAGYPSYSSAWMTLTQRBMGINEPKPDWWTYLDTA 389  
DB 421 MNGFVFVQSTDVRMPDPFNPFVKHEASYPHRSDAIWPLTQMRKGGQITESPASWYETA 480  
QY 390 KNVVRPDIYLAALAEELVAEGKAKADEPAD--TSIKSONFFIDKVPDPANKPNDYIAKF 447  
DB 481 KKIKRKPELYRAAALLISEGLDLPNELPAIDYDGRVATTEFIDGNKRYDAKDPIGYINSF 540

QY 448 AIGLKRGOTVA 458  
DB 541 EIGNKDDESILA 551

RESULT 6  
NRTC\_SYNY3 STANDARD; PRT; 670 AA.  
AC P73450;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Nitrate transport ATP-binding protein nrtc.  
Name=nrtc; OrderedLocNames=111452;  
OS Synecocystis sp. (strain PCC 6803).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.  
CX NCBI\_TaxId=1148;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97061201; PubMed=8905231;  
RA Kaneko T., Sato S., Korant H., Tanaka A., Asamizu E., Nakamura Y.,  
RA Miyajima N., Hitosawa M., Sugitara M., Sasamoto S., Kimura T.,  
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,  
RA Shimpou S., Takeuchi C., Wada T., Wakatane A., Yamada M., Yasuda M.,  
RA Tabata S.;  
RT "Sequence analysis of the genome of the unicellular cyanobacterium  
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the  
RT entire genome and assignment of potential protein-coding regions.";  
RL DNA Res. 3:109-136(1996).  
CC -!- FUNCTION: Probably part of a high-affinity binding-protein-  
CC dependent transport system for nitrate. Probably responsible for  
CC energy coupling to the transport system.  
CC -!- SUBCELLULAR LOCATION: Membrane-associated (Potential).



Query	Match	Score	DB	Length	441
Best Local Similarity	33.8%	Pred. No. 5,4e-38			
Matches	153	Conservative	74	Mismatches	182
				Indels	43
				Gaps	12
Qy	19 SLAVMGITLTPDVGAVGKLEKEDKFFGKLTMDAPLAAVAEKGFFFEDEGLFOLEAOAN	78			
Db	19 NLSAGGLPAP--GSIHESGRTYRAGFIFLVAAVAALIAAEFGFAOREGITLLELVDS	76			
Qy	79 WKVMDVAVNGEILDGSHMLAPAPLAASVGGGTADILEV--PFGNGFNATTVSNEIWHOM	137			
Db	77 WANRDLAFRQFVAMLSPMFPAAMLGGSNDSPPTIAPFSLDGGNATLTLRYGLM	136			
Qy	138 KPNPLEGKPVHPKADYLYKPVVEKKYKAEK--PFWAMTFPAGSHNIKRYWLAAGIN	196			
Db	137 QQAGGLGGGDDALK--NAKLLAAVIRSAAGAKKPLITGLVYPPSSHRYEERYWLAAGID	195			
Qy	197 PGYSPQPDIDISGQIGADALLSVTPPPQPMSTLEAGTIFGKCVGEPNNQOAVFKIGIYPI	256			
Db	196 P-----DREVKLVVVPPPTSDALAAAGIDGFCVGA PMWVVASEGVRIVA	242			
Qy	257 TDEELMDTPEKVFQGVTKQMAEKYPNTYLAVTALRAALWLDADNNKKNKEALIMLAOK	316			
Db	243 TKQDIMSAPAEKVIQMPPEWAEANRDVSRVLVALDPAAM--SDEPANGRLAEVIAEE	300			
Qy	317 QYVADAEVAVLAASNGTFEYEKD--DKRALPDFTFPHGASYSYSASAVWYLTQLRRGM	375			
Db	301 RHVAAPADDIRRYLAGEFALDPEGKRVYENVLYVHAGFANYRPSQALMTTYSQVMRMQ	360			
Qy	376 INEFKPDNMYLTPAKNYRPDIYLAAKELVAEGKAAEDPADTSI-----KPSONFI	430			
Db	361 TSLSRQ--RVDAALISAYRPDLVREAL-----GK--DALPADADAGRLEGATDGRFM	408			
Qy	431 DKVPFDANKENDYLAKFAIGLKKQYVAGSKV	462			
Db	409 DGHVFDGRIEDYIKSF-----GAPSVNSRV	435			
RESULT 10					
ID	08U924	PRELIMINARY	PRT	431	AA.
AC	08U924	07CTL2			
DT	01-JUN-2002	(TREMBlrel. 21, Created)			
DT	01-JUN-2002	(TREMBlrel. 21, Last sequence update)			
DT	25-OCT-2004	(TREMBlrel. 28, Last annotation update)			
DE	ABC transporter, nucleotide binding/ATPase protein (AGR_L1881g1p).				
GN	Name=rtc; OrderedLocNames=AGR_L1881g1, AUC3906;				
OS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;				
OC	Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.				
OX	NCBI_TaxID=16299;				
OX	NCBI_TaxID=16299;				
NP	[1]				
NP	SEQUENCE FROM N.A.				



RC STRAIN=Dupont;  
 RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;  
 RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.E., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Kap P.D., Bovee D.S.,  
 RA Chapman P., Clendinning J., Decherage G., Gilet W., Grant C.,  
 RA Kuyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Neister E.W.;  
 RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Cereon;  
 RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Guorollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,  
 RA Hounmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu P.,  
 RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT *Agrobacterium tumefaciens* C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AE009321; AA144714.1; -;  
 DR EMBL; AE008293; AA89513.1; -;  
 DR PIR; AD3037; AD3037.  
 DR PIR; G98248; G98248.  
 KW Complete proteome.  
 SO SEQUENCE 431 AA; 46254 MW; 22DB764D66992C63 CRC64;

Query Match 25.0%; Score 611.5; DB 2; Length 431;  
 Best Local Similarity 35.5%; Pred. No. 1.2e-37;  
 Matches 151; Conservative 69; Mismatches 166; Indels 39; Gaps 12;  
 DB 34 VGLKEKEDLKFGEFKITLDMAPLVAAEKGFEEDEGLFVQLEAQAANKVMDRVNDELQ 93  
 24 VSGDRKRIKAGFTPLVDASVLIAAEPGFADREGILDLVDKVSANVRDRLAERQFDI 83  
 94 SHMLAPAPLAASVGFSTKADIEV-PPSMGFNGNAITVSENIWOMKPNIP.L-EGKRPVAP 151  
 84 AHHLSMPVASMGLGNSNPSTPTTPPSLSRGNAITLSTRLPARMALGSLSTAGALEN 143  
 152 IKADYIKPVVEKYKAEK-G-PENNAWTFPAGSHNIKLRWYLAAGINPGYVSPQDISQOI 210  
 144 ARA--LKLVLDDMRARAGEAPPTLGMTYPFSSHNYEFERYWLAAGIHPDH----- 190  
 211 GADALLSVPPPMPTLEAGTIFGYCVGEPNQAVFVGIGVPIVTDLEIMKDPREKVF 270  
 191 --DVKLVVPPPTSDALAGALDGCVGAPMNIVAEKGVGIVAKKDLPESAEEKYI 248  
 271 GYTKQWAEKYPNTYLAFTVALLIPAAIWL-D-ADNNKRRKEAIEMLAQOYVGADEVYLAAS 329  
 249 GMRPEWABEQEIVGRLTLALDAASWCDLADHDLGSG---LADPRITGAPOSITRRV 305  
 330 MNGTEFEYE-KDDKRALPDENTFPRHGAASIPSSSAVWYLTQLRRWMINFEKDNMYLDT 388  
 306 LAGEFISIDSGNRRVIEKYFTFEGHDANYPROSQSLIYISQIMRWG---QAELSETGVNA 362  
 389 AKKVVPPDIYLAAKELVAEGKAKADPADPSIK-----PSQNFIDKTPFPAANKNDVL 444  
 363 ALSAVPDIYRAA-----LGDGKA-----PEDADIRIEGDEGDFVDFVPADTAGYV 413  
 445 AKFAI 449  
 414 NSPAV 418

RESULT 11

NRTA ANASP  
 ID NRTA ANASP STANDARD; PRT; 440 AA.  
 AC 044262; 006469;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 25-OCT-2004 (Rel. 45, Last annotation update)  
 DE Nitrate transport protein nrtA.  
 GN Name=nrtA; Ordered locus names=alr0608;  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97144534; PubMed=8990301;  
 RA Frias J.E., Flores E., Herrero A.;  
 RT "Nitrate assimilation gene cluster from the heterocyst-forming  
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL J. Bacteriol. 179:477-486(1997).  
 RN [2]  
 RP REVISIONS.  
 RA Frias J.E.;  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Matano A., Iriuguchi M., Ishikawa A., Kawashina K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 RN [4]  
 RP SEQUENCE OF 1-92 FROM N.A.  
 RX MEDLINE=97136629; PubMed=8982006;  
 RA Cai Y., Wolk C.P.;  
 RT "Nitrogen deprivation of *Anabaena* sp. strain PCC 7120 elicits rapid  
 RT activation of a gene cluster that is essential for uptake and  
 RT utilization of nitrate.";  
 RL J. Bacteriol. 179:258-266(1997).  
 CC -1- FUNCTION: Essential component of the nitrate-transporting system.  
 CC -1- May be the substrate-binding protein (By similarity).  
 CC -1- SUCCESFULR LOCATION: Inner membrane-associated (potential).  
 CC -1- SIMILARITY: Strong, to carotenoid-binding protein A (cbpA).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC  
 DR EMBL; X99709; CA68041.2; -;  
 DR EMBL; AP003583; BAB7566.1; -;  
 DR EMBL; U61496; AAC46075.1; -;  
 DR PIR; AG1882; AG1882.  
 DR InterPro; IPR006311; Tac.  
 DR TIGRFAMs; TIGR01728; Seta\_fam.1.  
 DR TIGRFAMs; TIGR01409; Tat\_signal\_seq.1.  
 KW Complete proteome; Inner membrane; Nitrate assimilation; Transport.  
 FT CONFLICT 100 W -> C (in Ref. 1).  
 SO SEQUENCE 440 AA; 48475 MW; 29937A41FB45CE9C CRC64;

Query Match 25.0%; Score 610.5; DB 1; Length 440;  
 Best Local Similarity 33.4%; Pred. No. 1.5e-37;  
 Matches 158; Conservative 71; Mismatches 173; Indels 71; Gaps 17;  
 DB 8 SSKLLT-----LSASLAWMLGTI-----APDVGANG--TLKEKEDLKF 44  
 5 SRRKPLFTTGAAASLIVHGTCTNSQSQTTEQAPSAAPANVSAANPVEVTTKAL 64



Oy		300	DDNNKRRKEALTEMLAQKVGVADVEVLAAASNGTGEY--EKDKRALDENTPFPHGASY	357
Dd		284	ADKANAKDELQVAVSRSSMIGAPVDIAVRKYGIIDYGGRVERNSPHINQFMWDFASY	343
Oy		358	PSYSAAVYLTLDRMGMINERKPPNNWLDTAK--NYYRPDIYAAAKELVAEKGAAAB	414
Dd		344	PYGSIDLWFPLETDIMGVLP-----TTDTGKLVAAYNRBDLMREAAKAL-----GQ	390
Oy		415	DEPADTSIKPSONFFIDKVPFEDANKENDYL	444
Dd		391	PAPKGTST-RGVKEFE-DGVAFDPPTXPEAVL	418
 RESULT 14 O8Y276 PRELIMINARY; PRT; 657 AA.				
ID	O8Y276			
AC	O8Y276;			
DT	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 25, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Nitrate transport ATP-binding protein.			
CN	Name=ntc; OrderedLocusNames=atrc0610;			
OS	Anabaena sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
OX	NCBI_Taxid=103690;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21595285; PubMed=11759840;			
RA	Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,			
RA	Matenabe A., Iriuguchi M., Ishikawa A., Kawashina K., Kimura T.,			
RA	Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,			
RA	Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,			
RA	Yasuda M., Tabata S.;			
RT	"Complete genomic sequence of the filamentous nitrogen-fixing			
RT	cyanobacterium Anabaena sp. strain PCC 7120.";			
RL	DNA Res. 8:205-213(2001).;			
CC	-1- SIMILARITY: Belongs to the ABC transporter family.			
CC	EMBL; AP003583; BAB72568.1; -.			
DR	PIR; A11882; A11882.			
DR	HSSP; O58206; 1L2T.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.			
DR	GO; GO:0015112; F:nitrate transporter activity; IEA.			
DR	GO; GO:0000166; F:nucleotide binding; IEA.			
DR	GO; GO:0015706; P:nitrate transport; IEA.			
DR	GO; GO:0006810; P:transport; IEA.			
DR	Pfam; PF00005; ABC_tran; 1.			
DR	Prodom; PD000006; ABC_transporter; 1.			
DR	SMART; SM00382; AAA; 1.			
DR	TIGRFAMS; TIGR01184; ntrCD; 1.			
DR	PROSITE; PS00211; ABC_TRANSPORTER_1; 1.			
DR	PROSITE; PS00893; ABC_TRANSPORTER_2; 1.			
KW	ATP-binding; Complete proteome; E7E1434798109324 CRC64;			
SQ	SEQUENCE 657 AA; 73294 MW; F7E1434798109324 CRC64;			
 Query Match 24.0%; Score 587; DB 2; Length 657; Best Local Similarity 35.8%; Pred. No. 1,5e-35; Matches 150; Conservative 74; Mismatches 139; Indels 56; Gaps 17				
Oy		39	KEDLKFGFIKLTMDAPLAVALAEKGFEEDEGL-FVLEQAQMWKVMDRVNNGELDGSHL	97
Dd		276	KAVIEIGEMPILTDSAPLIIVAKEKGFPAKGLDNVTILRNANMQALATAGTVVGKLDAAQNV	335
Oy		98	AAPAPLAASVGRTKADIEV--PFSNGPFGNALITVENETWHQMKNPIPLEGGKPVHPICAD	155
Dd		336	AGMPALTILTAGSQTTPPTVINALNISRNAMITTFSKRLYNQGVRSI-----AD	383
Oy		156	YLKPVEVKYKAEGKFPNMAMTFPAGSHNIKIERYMLAAGINPGVSPPODISGQIGADAL	215
Dd		384	LKAVIID--SEPDQLITLGCVASHASMOULIRTYMLAAGID-----PDRVVS-----	427
Oy		216	LSVTPPOMPSTLEAGTI FGVCVGPSPNQOAVFKGIGVPIVITDEBIMDPTREKVFQVTKQ	275

Db	Query Match	24.0%: Score 585.5; DB 2; Length 420; Best local Similarity 32.8%; Pred. No. 1.1e-35; Matches 151; Conservative 82; Mismatches 164; Indels 63; Gaps 15
Db	428 LTVLPTQWVSQLRAGNDIGYCAQSPMYQVAVHDLGFAVATALEIWSGQPKVGLGVRED	487
Qy	276 WAEKFPNTYLVATVTAIRAIWIMLADNNKRNKEAIEMLAQKQYGVADVEVLAASNGTF-	334
Db	468 WAQKYPETLYLVATLIEACKY--CDDLARNEEILEICREY-----LDVNPAYVRSGFI	541
Qy	335 -EYEKD--KRALPDENTPFRRHGAAPSYSAAVWYLTQLRRWGMINEFKPDNWTLDPAK	390
Db	542 DPYRGDGTPEPQOLTAAYNQFLINTNTYFNREIIMTIMQMARWG-LTPF-PKWN-VEITE	598
Qy	391 NVYPRDYLAAKELVAGKAKADEFPADTSIKSQNFIDKVPDPADKPKNDYLAKFAI	449
Db	599 RVCRDIDFGAARDL-----GLIDIGEDDP1---HLFDGL-FNPSEPLEYKLSLEI	646
RESULT 15		
OBY2F5		
ID	08Y2F5	PRELIMINARY; PRT; 420 AA.
AD	08Y2F5;	
DT	01-MAR-2002 (TREMBLrel. 20, Created)	
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)	
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)	
DE	PUTATIVE NITRATE TRANSPORTER PROTEIN.	
GN	Name=nstf; Synonyms=RS03349; OrderedAccession=RS0381;	
OS	Ralstonia solanacearum (Pseudomonas solanacearum);	
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;	
OC	Burkholderiaceae; Ralstonia.	
OX	NCBI_TaxID=305;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=GM1100;	
RX	MEDLINE=21681879; PubMed=11823852; DOI=10.1038/415497a;	
RA	Salanoubat M., Genin S., Artiguenave F., Guzy J., Mengnot S.,	
RA	Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,	
RA	Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,	
RA	Gaspin C., Levie W., Moisan A., Robert C., Saulin W., Schlex T.,	
RA	Stigler P., Thebault P., Whalen M., Winkler P., Levy M.,	
RA	Weissenbach J., Boucher C.A.;	
RL	"Genome sequence of the plant pathogen Ralstonia solanacearum.";	
RT	Nature 415:497-502(2002)	
DR	EMBL; AL646058; CADD3909.1; ..	
KW	Complete proteome.	
SC	SEQUENCE 420 AA; 45516 MW; 6BC0BDBABFB82F3C CRC64;	
Qy	3 TIIRSSSKKLTLTSS--ASLAWGLTTPADVGAVGK--LEKEDLKFGFIKLTDMAPLVA	58
Db	9 TPINPKRRRLVATYAAGSAGMALIDPLVRAGAMAAAGSDAPEKTELKVGRTIPITDCA	68
Qy	59 AEKGFEDDEGLFVLEAOAMKVMKRVNCGELDGSIMLAPLAAPLAASYGF-GTADIEVP	117
Db	69 ATLIGDKKYGIKIVPSKEASWAGVRDKLVSGDLAAHYLVGLVYGVOVIGIGPKKDMAVL	128
Qy	118 FSMFGNGAIVSNEIMHQMKPNPLBGGKRVHRIKADYLKPVVYKYAEGKPFNMATF	177
Db	129 MTLNNNGAIISSKL-----KAGVRDGAASLALMTR--EKRDYTFAGTF	172
Qy	178 PAGSHNITLRTWMLAAGINPGYSPPODISQIIGADALLSTVPEPOMSTLEAGTIFGVC	237
Db	173 PTGTHAMLVYTWLAHGHHP-----LQDAKAIIVPPPPQVANNRNVGMDQYC	219
Qy	238 VGEPMNQDAVEFKIGVPIVITDELMLKDTPEKVFSGVYKQWAEKYPNTYLAVTKALIRAIW	297
Db	220 VGEPMQARAIDAGIGFAETFGATIMKDHPEKVLGTAAEFQKYPNTARALTAAYLAASKF	279
Qy	298 LDADNNKRRKKAIEMLAQKQYGVADVEVLAASNGTFP-----EKDDGRALPDPTF	350
Db	280 IDA-SASNRRTAETVTAASVYNTDMDIILDRLGRYTNGLGKTWDDADPWR-----F	331

Qy 351 FRHGA-SYPESSAVWYLTOLRRWGMINEFKPDNWTLDTAQNYYRPDIYLAAKEIVABG 409  
Db 332 YHDAVWPFYLDGMMWFLIOHKRWGL-KTHPD-YLAIAQVNRVDIYKQAA----- 381  
Qy 410 KAKAEDFPADTSIKPSQNFIDKVPDPANKPNDYLAKFAI 449  
Db 382 AATGTPLP-KSDLRTAR--LIDGVWMDAKNPAAYADSFKI 418

Search completed: September 16, 2005, 01:09:11  
Job time : 181 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 00:38:09 ; Search time 44 Seconds

(without alignments)  
787,208 Million cell updates/sec

Title: US-10-689-200-2

Sequence: 1 MKTIRSSSKKLLTSLASU.....AKFAIGKQIVAGKVD 464

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/aa/6A\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/aa/6B\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/aa/6C\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/aa/6D\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/aa/6E\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	577.5	23.6	456	US-09-489-039A-10363	Sequence 10363, A
2	446.5	18.3	474	US-09-540-236-2892	Sequence 2892, Ap
3	338.5	13.9	336	US-09-328-352-7572	Sequence 7572, Ap
4	292.5	12.0	249	US-09-252-991A-30501	Sequence 30501, A
5	240	9.8	419	US-09-252-991A-25707	Sequence 25707, A
6	219.5	9.0	275	US-09-602-787A-384	Sequence 384, App
7	157	6.4	197	US-09-252-991A-30392	Sequence 30392, A
8	121.5	5.0	336	US-09-902-540-12264	Sequence 12264, A
9	116.5	4.8	623	US-09-252-991A-26757	Sequence 26757, A
10	116.5	4.8	337	US-09-252-991A-25810	Sequence 25810, A
11	103.5	4.2	327	US-09-489-039A-12136	Sequence 12136, A
12	102	4.2	369	US-09-107-532A-5754	Sequence 5754, Ap
13	102	4.2	1062	US-09-397-550-4	Sequence 4, App1
14	102	4.2	1076	US-09-470-443-6	Sequence 6, App1
15	102	4.2	1082	US-09-397-550-5	Sequence 5, App1
16	102	4.2	1109	US-09-397-550-6	Sequence 6, App1
17	102	4.2	1115	US-09-397-550-23	Sequence 23, App1
18	102	4.2	1145	US-09-470-443-2	Sequence 2, App1
19	102	4.2	1145	US-09-470-443-4	Sequence 4, App1
20	102	4.2	1145	US-09-397-550-20	Sequence 20, App1
21	101.5	4.2	952	US-09-328-352-5611	Sequence 5611, Ap
22	101.5	4.2	1252	US-10-012-762-20	Sequence 20, App1
23	101.5	4.2	1252	US-09-704-036B-20	Sequence 20, App1
24	101	4.1	858	US-09-255-829-22	Sequence 22, App1
25	101	4.1	858	US-09-255-829-29	Sequence 29, App1
26	101	4.1	1169	US-09-255-829-20	Sequence 20, App1
27	98.5	4.0	315	US-09-248-796A-16091	Sequence 16091, A

28	98.5	4.0	1864	US-08-804-227C-3	Sequence 3, App1
29	98	4.0	324	US-08-956-171E-5209	Sequence 5209, Ap
30	98	4.0	324	US-08-781-986A-5209	Sequence 5209, Ap
31	96.5	4.0	448	US-09-198-452A-216	Sequence 216, App
32	96.5	4.0	457	US-09-438-185A-199	Sequence 199, App
33	96.5	4.0	659	US-08-258-639A-4	Sequence 4, App1
34	96.5	4.0	659	US-08-900-951-4	Sequence 4, App1
35	96.5	4.0	659	PCT-US95-07391A-4	Sequence 4, App1
36	95.5	3.9	893	US-09-489-039A-14127	Sequence 14127, A
37	93.5	3.8	540	US-09-302-540-14003	Sequence 14003, A
38	93.5	3.8	1018	US-08-452-052-2	Sequence 2, App1
39	92	3.8	620	US-09-442-100-11	Sequence 11, App1
40	92	3.8	620	US-08-939-105-11	Sequence 11, App1
41	92	3.8	620	US-09-442-102-11	Sequence 11, App1
42	91	3.7	379	US-09-071-035-6	Sequence 6, App1
43	91	3.7	402	US-09-489-039A-11633	Sequence 11633, A
44	91	3.7	431	US-09-134-000C-6307	Sequence 6307, Ap
45	91	3.7	777	US-09-917-254-57	Sequence 57, App1

#### ALIGNMENTS

RESULT 1									
US-09-489-039A-10363									
Sequence 10363 Application US/09489039A									
Patent No. 6610836									
GENERAL INFORMATION:									
APPLICANT: Gary Breton et. al									
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA									
TITLE OF INVENTION: PNEUMONIA FOR DIAGNOSTICS AND THERAPEUTICS									
FILE REFERENCE: 2709.2004001									
CURRENT APPLICATION NUMBER: US/09/489,039A									
CURRENT FILING DATE: 2000-01-27									
PRIOR APPLICATION NUMBER: US 60/117,747									
PRIOR FILING DATE: 1999-01-29									
NUMBER OF SEQ ID NOS: 14342									
SEQ ID NO 10363									
LENGTH: 456									
TYPE: PRT									
ORGANISM: Klebsiella pneumoniae									
US-09-489-039A-10363									
Query Match									
Best Local Similarity 23.6%; Score 577.5; DB 4; Length 456;									
Matches 153; Conservative 76; Mismatches 173; Indels 55; Gaps 17;									
QY	7	SSSKKLLTSLASIAVGLTIAIPDV-----GAVGKLEKEDLKFGFKLTDPAPLVAAB	60						
DB	44	STSRRLDGAALG--GAMLLPGVQAAAGSGDKPEQTVRVGFTPLTDCAPLAIYAA	101						
QY	61	KGFEEDEGLFVQLEAQAANKVMDRVVNGELDGSNMLAPLAASVFGTKADIEVP-PS	119						
DB	102	KGGDQKYGTTIVAAKSAKSAWAARDKLVAGELDAHLVGLVGLGELGIAKSPGAMNMT	161						
QY	120	MENGNATVSNIEIWMKPNIPLEGKRVHPKADYLPVVEKRYAABGKPEMNANTFPA	179						
DB	162	LNNNGCAITLSESL--QEGKGVTDLG-----LKRLLDR-SAPSS-YFPAHTFPT	206						
QY	180	GSHNITLRLRWLAAGINPGYSPPODISGIGADALLSTPPROMSTLEAGTIFCYCG	239						
DB	207	GTHAWMLYTWLSAGIDP-----FNDVTVVPPQVMNRKIGMNSGFCVG	253						
QY	240	EPNNOQAVFKIGVPIYDDELKMDTPEKVFQVTKQMAEKYPTVYLAATKALIRAIWLD	299						
DB	254	EPNNAIINDRIGFTAATSGQDIWPEHEKVLGTRBRDVRNNTAALVAALMEARWT-	312						
QY	300	ADNNKQKKAILEMAQKQYVGADEVYLAASNGTPEYEDDKRALPDFN--TFRRIG-AS	356						
DB	313	AASPEHTRETAIRLARNGWLNTRKEQYLTGRMG--EYDNGLRBWDAPHRIFRWAGS	370						
QY	357	YFSYSAVWYLTQLRWGMINEKPKPNWYLDTPAKVYRDIYLAALKEIVAGSKAAEDF	416						

Db 371 FFWLSDGWMFLTQFRRMGLLKQ-APD--YLAVASRINRIDVMOAAQ---AVGISA---- 421  
QY 417 PADSTIKPSQNFIDKVPDANKPNDYLAKEAIGKG 453  
Db 422 PA--ARMRSITLMDGTVMNGSDPEGARHFSIORKG 455

## RESULT 2

US-09-540-236-2892  
; Sequence 2892, Application US/09540236  
; Patent No. 6673910  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CAVAT  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2005-001  
; CURRENT APPLICATION NUMBER: US/09/540,236  
; NUMBER OF SEQ ID NOS: 3840  
; SEQ ID NO 2892  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: M.catarhalis  
US-09-540-236-2892

Query Match 18.3%; Score 446.5; DB 4; Length 474;  
Best Local Similarity 27.9%; Pred. No. 4.5e-36;

Matches 199; Conservative 82; Mismatches 180; Indels 71; Gaps 11;

QY 2 KTIIRSSSKLLTLTSLASLAWGLTIAPDVGAVGKL--EKEDLKFGFIKLTDMAPLAVAA 59  
Db 79 KTVGGTTA-----LAAISVLEPIATLQEAALIDTLKPEKSSVDIGFILTATPLIMAD 132  
QY 60 EKGFFEDGELFVQLEQAAMKVMDRVNGELDGSMTLAPAPLAASVGGT--KADIEVFP 118  
Db 133 PLGYVAEOGIKANLLKRAAGALVRDMMNRRELDAAHFLAPMLAINLGLGSAKONKAVAA 192  
QY 119 SMGFNGNAITVSNEIWMQKNPIPLEGKVPHPKADYLPKPVVEKYKAEGKPFNMAMTFP 178  
Db 193 IONTNGQALVMA-----LKHKNRNPKNPKWKMTFAPLPE 226  
QY 179 AGSHNIKRLRYWLAAGINPGYVSPPODISGQIGADALLSVTPPQMPSTLEAGTIFGCV 238  
Db 227 HSHINLRLRYFLAEHGLD-----DKDVKLRLTPRPDIAMILKAGNIDGFG 273  
QY 239 GEPNNOAIFKGIQVIVIDEIMKDTPEKVGVTQMAEKYQNTYLAITKALIRAIYL 298  
Db 274 PEPFNORAVWDKAGYIHTLSRDIWNGHPCCSFGTSQSFINDYPPQTLAVYRAIILKANVM- 332  
QY 299 DADNNKRRKEATLMLAQOQYGVADVEVLAASNMGTFEYEKDDKRALPDFTFRRHGA-S-Y 357  
Db 333 -ANKPSIRKDLSTLSPAYLNPBLVLOSINGRRLADGVDTIGQVDP-----RKGFDM 386  
QY 358 PSYSSAVWYLTQLRMGMINFEKPDNWDYDTAKNVYRPDIYLAALAEVAEGKAKADEFP 417  
Db 387 PMSVAAMMMTQMKRGYIT---GNINYODIANQVE---MLTAKKQOMAGYTVADDBP 440  
QY 418 ADTSIKPSQNFIDKVPDANKPNDYLAKEATLKGKQYVAG 459  
Db 441 KITVWGKQ-----FNATPDAYLDSFALIGHKSTGRILHG 473

## RESULT 3

US-09-328-352-7572  
; Sequence 7572, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7572  
; LENGTH: 336  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7572

Query Match 13.9%; Score 338.5; DB 4; Length 336;  
Best Local Similarity 26.5%; Pred. No. 2e-25;

Matches 96; Conservative 63; Mismatches 144; Indels 59; Gaps 9;

QY 34 VKLEKEDLKFGFIKLTDMAPLAVAAEKGFEDGELFVQLEQAAMKVMDRVNGELDQ 93  
Db 1 MSKLEKTOQDQYITFLDLOIALMLMAKQGFEEVDLDVTLYEASWASIRDLARGLDGA 60  
QY 94 SMTLAPAPLAASVGGTAD-----IEVPSGFGNNAITVSNELIWMQ---KRNIPLE 144  
Db 61 AHCLSAMLPAAAMG---ADQIGALQPLVLSKRRAPFISLSQKLIHOLAIRENDNAQTT 116  
QY 145 GGGPVHPKADYLPKPVVEKYKAEGKPFNMAMTFPAGSHNIKRLRYWLAAGINPGYVSPQ 204  
Db 117 AOKVIOYIEODH-----TSLAHVFKHSIHYYCLREWLALA----- 152  
QY 205 DISGQIGADALLSVTPPQMPSTLEAGTIFGCVGEPNNOAIFKGIQVIVIDEIMKMD 264  
Db 153 --DSRIAGTLKALKALPPRYWTEALDNHVIDGFCVGEPMNTQGEILGLSKIVCSQDIIEN 210  
QY 265 TREKYFGVTQMAEKYQNTYLAITKALIRAIWLDADNNKRRKEATLMLAQ---KQYVG 320  
Db 211 VADKILAVTQEWAEQHPTLVALLTAIMKAQ--KELSYLKPAPALYKLVFEGIVRFHCS 268  
QY 321 ADVEVLAASNMGTFEYEKDDKRALPDFTFRRHGA-SYSSAVWYLTQLRMGMINPEX 380  
Db 269 EEVNDKTYMIONIVKHLVKENAARQEDFH-----WLFQOMQMKELQ-LQA 314  
QY 381 PD 382  
Db 315 PD 316

## RESULT 4

US-09-252-991A-30501  
; Sequence 30501, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUSINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30501  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30501

Query Match 12.0%; Score 292.5; DB 4; Length 249;  
Best Local Similarity 31.3%; Pred. No. 5.4e-21;

Matches 73; Conservative 40; Mismatches 105; Indels 15; Gaps 6;

QY 218 VTPPQMPSTLEAGTIFGCVGEPNNOAIFKGIQVIVIDEIMKDTPEKFGVTQMA 277  
Db 19 VVPRQMGHIOAGRIDFCAGCPWGAALAVDQGGFTATISGAIYRDPHEKLTGTRAAV 78  
QY 278 EKYPTTYLAITKALIRAIWLDADNNKRRKEATLMLAQOQYGVADVEVLAASNMGTFEYE 337  
Db 79 DAYPNTABALVWAVLDASRFTL--QNAENRLGTAQLISGRDYVADALGAIQPRFFGRYQDG 137



```

QY 338 KODKRALPDENTFRHG-ASYPSYSSAVWLTOLRRMGMINERKPDNWTIDTAKNYRDP 396
DB 138 LGNAWODPHRLRYADGEVVRPMLPSGMFMFTQFRRWGLRE--DPP--YLGARRVQOQA 194
QY 397 IYLAAKEVVAEGKAKADEFPADTSIKPSQNFIDKVPDPANKPNYLAKFAI 449
DB 195 LVKDAATLGL--RLDGADYKRST-----LIDRTWDSDPAGYASFP 237

RESULT 5
US-09-252-991A-25707
; Sequence 25707, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25707
; LENGTH: 419
; TYPE: PR
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25707

Query Match 9.8%; Score 240; DB 4; Length 419;
Best Local Similarity 24.6%; Pred. No. 2.5e-15;
Matches 93; Conservative 68; Mismatches 141; Indels 76; Gaps 14;

QY 9 SKKLLTLSASLAVNG-LTIAPDVGAVKLEKE-DKPGRIKLTDMAPLVAAEKGFED 66
DB 29 SRRDIKTLALISAAGALPLSLQRAAEPAPRIGLPTDTPPLLVANGLFPA 88
QY 67 EGLFVQLEAQA--NNKVVMDRVVNGELDGSMLAPAPLASVGFSTKADIEVPSMGFN 123
DB 89 EG--IOAERPVLLRSNAQVIEAFISQGVNVHLSMTYMAR--YSSKVPATVANNHNG 144
QY 124 GNAITVSNIEIHWQMKENIPLEGKPVHPITKADYLKPEVKEAEGKPFNMAMTFPAGSHN 183
DB 145 GSGLTVAPEI-----ADVRLGGKSV-----AIPFVYSIHN 175
QY 184 IKLRVYLAAGINPGYSPPODISGOIGA-----DALLSVTPPPQMPSTLEAGTIFGCY 238
DB 176 VVLQQLLRDNGL-----RAVSRVGAALADEVNLVLPSPDMPPLASRRIHGYIV 227
QY 239 GEPANQAAVFKGIVETIDEELMKOTPKVFGVTQMAEKYNTYLAATKALIRAAIWL 298
DB 228 AEPFNAALANLKVGRVQRTGDMWRHACCVFMHEDERRPQWQKYNVAIVKQQLM-- 286
QY 299 DADNNKREKALEMTAQ--KQYGVADVEVL-----AASMGSTFEYKQDKXA 343
DB 287 ---TEHRAEAQAQLSKAGANRTPHAPVYLGVLAPGAEEQAAVLASAIRHADQGER 343
QY 344 LPDENTFRHGASYPSS 361
DB 344 I-DFOPY-----PYPSYT 355

RESULT 6
US-09-602-787A-384
; Sequence 384, Application US/09602787A
; Patent No. 6696561
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Mark
; APPLICANT: Krüger, Burkhard
; APPLICANT: Schödel, Hartwig

```

```

; APPLICANT: Zelder, Oskar
; APPLICANT: Habener, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-125CP
; CURRENT APPLICATION NUMBER: US/09/602,787A
; CURRENT FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USSN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932182.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932190.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932191.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932209.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932212.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932227.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932228.7
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932229.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932230.9
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932927.3
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933005.0
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19933006.9
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19940764.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940765.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940766.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940830.0
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940831.9
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940832.7
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19940833.5
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19941378.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19941395.9
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19942077.7
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: DE 19942078.5

```

## RESULT 8

Query Match 4.8%; Score 116.5; DB 4; Length 337;

[illegible]

```

QY 326 -----LAASNMGTEVEYEDKRALPDENFTF-----FRHGASVPSYS 362
| | | | |
DB 449 AMKEPEYRALATSID-TLKLKKADRG-L-DVERFVDDKRIREARYQAQGLDYQALADYAP 506
| | | | |
QY 363 AWMYLTQLRRWGMINEFK--PDNMYLDTA--KNVYRPDIYLAARELVAEKRA---KAE 414
| | | | |
DB 507 SPLRASDARSGRPIEFERVAQIWKGEARVHYASPESALADLDELKQGAIRAIYAO 566
| | | | |
QY 415 DFPADTSIKPSQNFPI 430
| | | | |
DB 567 DRDSGIKILANQAMFV 582
| | | | |

RESULT 11
US-09-489-039A-12136
; Sequence 12136, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489.039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12136
; LENGTH: 327
; TYPE: PRF
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12136

Query Match 4.2%; Score 103.5; DB 4; Length 327;
Best Local Similarity 20.3%; Pred. No. 0.099;
Matches 65; Conservative 53; Mismatches 137; Indels 65; Gaps 12;

QY 19 SLAWGTLIADVAGVGLKEKDLKFGYIKLTDAPLVAAREKGFEEDEGLFVQLEAQN 78
| | | | |
DB 14 ALTTGGLLFSSLGHAPADAPDALRTIGYQKGSVSKVLA-----KSHQLTEORYPOTH 65
| | | | |
QY 79 WKVWMDRVNGELDGSHTLAPAPLAASVFGTKADIEVPFGMGFNAGAITVSENIWHQK 138
| | | | |
DB 66 ISWI-----EPFAGPQMLEALNWGSDIDISTGDIPIPIFQAAGADLVYGAE---PPX 115
| | | | |
QY 139 PN---IPLEGGKPVPIKADYIKPVEKEYKAEGKPFNNAMTPPAGSHNIKRLRYWLAAG-- 193
| | | | |
DB 116 PKAEVILVAQSGPIHNV-----TELKGGKYVAFQKSSSHNLIRALQLAGLK 162
| | | | |
QY 194 --GINPGYSPPOISQIGADALLSVTPPQMBSITLBAHTIFGCVGEPPNNQAVFQGI 251
| | | | |
DB 163 FSDIQPYVLA-----ADA-----RAAFQGGVDAVAIWDPPYSALLQG-- 202
| | | | |
QY 252 GVPIITTEELTKDTPKEKFGVTKQMAEKYPNTYLAVTALIRAIWTLADNNKKEKALE 311
| | | | |
DB 203 GARVLITGTDLQKTI-GSFYLAASRYAERNG---AFIEGVLDFTFGADALTHSQRAQSIT 257
| | | | |
QY 312 MLAGQYVADDEVYLAASMN 331
| | | | |
DB 258 LIAKTI--MGLPEAVIAYSLD 275
| | | | |

RESULT 12
US-09-107-532A-5754
; Sequence 5754, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:

```

```

/ ADDRESS: GENOME THERAPEUTICS CORPORATION
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD-ROM ISO9660
/ COMPUTER: PC
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: ASCII
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,532A
/ FILING DATE: 30-Jun-1998
/
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 5754:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 369 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: YES
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/
/ NAME/KEY: m1sc.feature
/ LOCATION: (8) LOCATION 1...369
/ SEQUENCE DESCRIPTION: SEQ ID NO: 5754:
US-09-107-532A-5754

Query Match      4.2%; Score 102; DB 4; Length 369;
Best Local Similarity 21.5%; Pred. No. 0.17;
Matches 85; Conservative 47; Mismatches 138; Indels 126; Gaps 20;

QY 10 KTLILLTSLASLAWGLTIAPDVGA---VGKLEKEDLKFGFIKLTMDAPLAVAEKGFED 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 41 KKLWLLPLLLLSACGTAKETSSKQEIKDCLKVTLLVDYVPNTNHTGYIYLAKEGYRKE 100
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 EGLFVOL-----EAQANKVVDVRYVNGELDGSMLAPAPLAASY----- 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 101 AGLVNQTIERGDNSTISGLVGADKQFGVSYQED-VTYAHADQGN--EVKATATYIKIN 157
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 107 --GFGTKADIEVPFSGNGFNALITYVNEIMHQMKPNIPLEGKRVHPRIKADYLKPVVEKY 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 158 TSGFATLSDSNHISPDGFKTYAG---WQS-----PSEEAVLKAVMERK- 198
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 165 KAEGRFNNAMTFPPAGSHNIIKRYWLAAGINPGYSPPODISGQ--GADALLSVTPPQ 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 --DGGDFS-----KLT-MVGSNGEGESIGKSSDIQWYEGWD----- 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 224 MSTLEAGTIFGCVGEPMNQAVFKGIGVPI-ITDEILMKDTPPEKVFVTKQMAEKYPN 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 234 MIKAKAGAEIENVYIPKEIDERLDY---TPVITTDQLIKSDPELV-----Q 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 283 TYLAVTKALIRAIWLDADNNKRRKEAIEMLAQKQYGVADVEVLASNGTPEYEKDDKR 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 SFWDATKKGYQEA---KDPNDSAK-----LLOK-----YAKENDR 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 ALPDFNTFPRHGAASYPSYSAWYILTQLRRMGIN 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 -----TFLEESQAFLSKN-----YTDDPRNWGIMER 337
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
US-09-397-550-4
/ Sequence 4, Application US/09397550
/ Patent No. 6783952
/ GENERAL INFORMATION:
/ APPLICANT: Warner-Lambert
/ TITLE OF INVENTION: Secreted soluble alpha2 delta-2, alpha2 delta-3, alpha2
/ TITLE OF INVENTION: delta-4 calcium channel subunit polypeptides and
/ TITLE OF INVENTION: screening assays using same
/ FILE REFERENCE: 180
/ CURRENT APPLICATION NUMBER: US/09/397,550
/ CURRENT FILING DATE: 1999-09-16
/ NUMBER OF SEQ ID NOS: 28
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 4
/ LENGTH: 1062
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-397-550-4

Query Match      4.2%; Score 102; DB 4; Length 1062;
Best Local Similarity 23.7%; Pred. No. 0.95;
Matches 79; Conservative 42; Mismatches 142; Indels 70; Gaps 17;

QY 7 SSKKLLTSLASLAWGLTIAPDVGAVGKLEKEDLKFGFIKLTMDAPLAVAEKGFED 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 287 SSPKDMVIVDVSSVSGSLTL-----KLMKTSVCEMLDPTLSDDDYVNVASFPKQAP 338
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 EGLFVOL-EAQANKVVDVRYVNGELDGSMLAPAPLAASYVGFGRKADIEVPFSGNGFN 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 339 VSCFTHLVQANVRNKKVFKEAVQ-----MVAKGITGYKAGFEYAFDQLQNSN 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 126 AITVSNEMIMHQMKPNIP--EGKRVHPRIKADYLKPVVEKYKAEKPFNNMAMTFPAGSHN 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 -ITRAN-----CNKXIMMFTDGE-----DRVDVEKYNWPNRTVRV-FTFSVGQHN 432
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 IKLR--YWLAAGINPGYSPPODISG-QIGADALLSVTPPOMSTLEAGTIFGVCVGE 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 433 YDVTPLQMMACA--NKGYTFEIPISGAIKIRINTQETLDVGRPMVLAGEAKQY----- 483
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 241 PWNQAVRK-GIGVEVITDEILMKDTPPEKVFVTKQMAEKYPNTYLAVTKALIRAIWLD 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 484 QWTN--VYEDALGLGLVVTGTL-----PFENLTQDGRGEKKN-----QLILGVMGID 528
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 300 ADNNKRRKEAIEMLAQKQYGVADVEVLASNG 332
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 529 VALND-----IKRLTPNTYTLGANGYVFAIDING 556
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-470-443-6
/ Sequence 6, Application US/09470443
/ Patent No. 6441156
/ GENERAL INFORMATION:
/ APPLICANT: Lerman, Michael I.
/ APPLICANT: Minna, John D.
/ APPLICANT: Latif, Farida
/ APPLICANT: Wei, Ming-Hui
/ APPLICANT: Sekido, Yoshitaka
/ APPLICANT: Gao, Boning
/ APPLICANT: Duh, Fuh-Mei
/ TITLE OF INVENTION: Calcium Channel Compositions and Methods of Use Thereof
/ FILE REFERENCE: NIH-05043
/ CURRENT APPLICATION NUMBER: US/09/470,443
/ CURRENT FILING DATE: 1999-12-22
/ EARLIER APPLICATION NUMBER: 60/114,359
/ EARLIER FILING DATE: 1998-12-30
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 1076
/ TYPE: PRT

```



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 16, 2005, 00:47:41 ; Search time 172 Seconds  
(without alignments)  
1092.363 Million cell updates/sec

Title: US-10-689-200-2  
Perfect score: 2443  
Sequence: 1 MKTIRSSSKLLTSLASL.....AKFAIGKKGQVAGKAVD 464

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1812044 seqs, 404927589 residues

Total number of hits satisfying chosen parameters: 1812044

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10E\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
21: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
22: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2443	100.0	464	US-10-689-200-2	Sequence 2, Appl1
2	525	21.5	403	US-10-282-122A-69975	Sequence 69975, A
3	240	9.8	399	US-10-282-122A-66370	Sequence 66370, A
4	240	9.8	399	US-10-389-647-502	Sequence 502, App
5	234.5	9.6	396	US-10-282-122A-49897	Sequence 49897, A
6	227	9.3	257	US-10-282-122A-44527	Sequence 44527, A
7	226	9.3	294	US-09-738-626-4906	Sequence 4906, Ap
8	222.5	9.1	372	US-10-282-122A-68363	Sequence 68363, A
9	220.5	9.0	391	US-10-282-122A-49957	Sequence 49957, A
10	219.5	9.0	275	US-10-627-476-384	Sequence 384, App
11	217.5	8.9	391	US-10-282-122A-48182	Sequence 48182, A

12	144	5.9	342	15	US-10-282-122A-72603	Sequence 72603, A
13	140	5.7	192	15	US-10-282-122A-47676	Sequence 47676, A
14	116.5	4.8	883	14	US-10-156-761-8511	Sequence 8511, Ap
15	113	4.6	1355	18	US-10-450-763-36085	Sequence 36085, A
16	111.5	4.6	353	14	US-10-156-761-9969	Sequence 9969, Ap
17	111	4.5	1357	18	US-10-450-763-39994	Sequence 39994, A
18	108	4.4	533	15	US-10-425-114-51223	Sequence 51223, A
19	107	4.4	518	15	US-10-424-559-198653	Sequence 198653, A
20	106.5	4.4	461	15	US-10-282-122A-46659	Sequence 46659, A
21	105	4.3	419	15	US-10-282-122A-46437	Sequence 46437, A
22	105	4.3	471	15	US-10-425-115-185585	Sequence 185585, A
23	105	4.3	1390	15	US-10-275-595A-6	Sequence 6, Appl1
24	105	4.3	3352	14	US-10-156-761-7961	Sequence 7961, Ap
25	105	4.3	4455	15	US-10-287-228-304	Sequence 304, App
26	104	4.3	866	14	US-10-241-556-104	Sequence 104, App
27	104	4.3	1420	14	US-10-241-556-110	Sequence 110, App
28	103.5	4.2	280	9	US-09-815-242-11714	Sequence 11714, A
29	103	4.2	864	14	US-10-241-556-102	Sequence 102, App
30	103	4.2	1139	16	US-10-417-375-176	Sequence 176, App
31	103	4.2	1141	16	US-10-417-375-174	Sequence 174, App
32	103	4.2	1142	16	US-10-417-375-172	Sequence 172, App
33	102.5	4.2	320	15	US-10-282-122A-55774	Sequence 55774, A
34	102	4.2	613	15	US-10-369-493-18239	Sequence 18239, A
35	102	4.2	860	14	US-10-241-556-175	Sequence 175, App
36	102	4.2	862	14	US-10-241-556-94	Sequence 94, Appl
37	102	4.2	866	14	US-10-241-556-88	Sequence 88, Appl
38	102	4.2	867	14	US-10-241-556-96	Sequence 96, Appl
39	102	4.2	867	14	US-10-241-556-98	Sequence 98, Appl
40	102	4.2	870	14	US-10-241-556-92	Sequence 92, Appl
41	102	4.2	871	14	US-10-241-556-84	Sequence 84, Appl
42	102	4.2	871	14	US-10-241-556-86	Sequence 86, Appl
43	102	4.2	871	14	US-10-241-556-90	Sequence 90, Appl
44	102	4.2	1062	17	US-10-902-531-4	Sequence 4, Appl1
45	102	4.2	1076	14	US-10-116-949-6	Sequence 6, Appl1

## ALIGNMENTS

RESULT 1  
US-10-689-200-2  
Sequence 2, Application US/10689200  
Publication No. US20040126848A1  
GENERAL INFORMATION:  
APPLICANT: Dicostimo, Deana J.  
APPLICANT: Ni, Hao  
APPLICANT: Ye, Rick  
APPLICANT: Picataggio, Stephen  
APPLICANT: Wang, Tao  
APPLICANT: Seip, John E.  
TITLE OF INVENTION: NATURAL PROMOTERS FOR GENE EXPRESSION IN C1 METABOLIZING BACTERI  
FILE REFERENCE: C1147 US NA  
CURRENT APPLICATION NUMBER: US/10/689,200  
PRIOR FILING DATE: 2003-10-20  
PRIOR APPLICATION NUMBER: 60/419,872  
PRIOR FILING DATE: 2002-10-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 2  
LENGTH: 464  
TYPE: PRT  
ORGANISM: Methylobacillus sp. 16a  
US-10-689-200-2

Query Match 100.0% ; Score 2443 ; DB 16 ; Length 464 ;  
Best Local Similarity 100.0% ; Pred. No. 3.2e-216 ;  
Matches 464 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;

QY 1 MKTIRSSSKLLTSLASLAWGTLTAPDVAGVKEKEDEKFGIKLTQMAPLAAVAAE 60  
DB 1 MKTIRSSSKLLTSLASLAWGTLTAPDVAGVAGVKEKEDEKFGIKLTQMAPLAAVAAE 60  
QY 61 KGFEDDEGLFVQLEAQAQWKKVMDRVNGELDGSHMLAPAPLAAVSGFGTADIEVPSFM 120



```

Db      61 KGFEEDEGLFVQLAEQAANKVMDRVNGLDGSHTLADAPLAASVGFCTKADIEVPSFM 120
Qy      121 GNGNAAITVSNETIHWQMKNPITLEGKPVHPKADYLKVVKKYAEKGFPMNMTFPG 180
Db      121 GNGNAAITVSNETIHWQMKNPITLEGKPVHPKADYLKVVKKYAEKGFPMNMTFPG 180
Qy      181 SHNIKRLRYLLAAGGIPGYSPPODISGQIGADALSVTPPOMPSTLEAGTIFGCYGE 240
Db      181 SHNIKRLRYLLAAGGIPGYSPPODISGQIGADALSVTPPOMPSTLEAGTIFGCYGE 240
Qy      241 PNNQOAVFVGIGVPTTDELMKDPTEKVFVYTKQMAEKYPTVYLAVTKALIRAAIWLDA 300
Db      241 PNNQOAVFVGIGVPTTDELMKDPTEKVFVYTKQMAEKYPTVYLAVTKALIRAAIWLDA 300
Qy      301 DNNKRRKEAIEMLAQOQYGVADVEVLAASNGTFFEEKDKRALPNTFFPHGASYSBY 360
Db      301 DNNKRRKEAIEMLAQOQYGVADVEVLAASNGTFFEEKDKRALPNTFFPHGASYSBY 360
Qy      361 SSAVWYLTQLRRGMINEFKPDNMYLDTAKNYRPTIYLAALKELVAEGKAAEDFPADT 420
Db      361 SSAVWYLTQLRRGMINEFKPDNMYLDTAKNYRPTIYLAALKELVAEGKAAEDFPADT 420
Qy      421 SIKPSQNFPTDKVPFDANKPNDYLAKEFALGLKQKQTVAGAKVVD 464
Db      421 SIKPSQNFPTDKVPFDANKPNDYLAKEFALGLKQKQTVAGAKVVD 464

```

```

RESULT 2
US-10-282-122A-69975
; Sequence 69975, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69975

```

```

; LENGTH: 403
; TYPE: PRT
; ORGANISM: Pseudomonas syringae
US-10-282-122A-69975

```

```

Query Match          21.5%; Score 525; DB 15; Length 403;
Best Local Similarity 31.7%; Pred. No. 2,5e-39;
Matches 132; Conservative 72; Mismatches 164; Indels 48; Gaps 12;

```

```

Qy      38 EKEDLKRFPILTDMAPLVAAEKGFEEDEGLFVQLAEQAANKVMDRVNGLDGSHTL 97
Db      19 EKNSLDVGFMALTDAPLVVAATQGFAPYGIUSLTKQTSWAGLRDLVSGQLAAHSL 78
Qy      98 AAPALAASVGF--GTRKADIEVPSFMSGFNGNAITVSNETIHWQMKNPITLEGKPVHPKAD 155
Db      79 YGLIYAVELGISGGPATTMALIMGINQGCCINLSRF-----LQDAGVITPEALD 128
Qy      156 YLKPVEKRYKAEKGFPMNMTFPAGSHNIKRLRYLLAAGGIPGYSPPODISGQIGADAL 215
Db      129 -----KQAHQSGSKLTFQAQTFPTGNHAMWLYWMLASQIH-----LDDV 168
Qy      216 LS-VTPPOMPSTLEAGTIFGCYGEFPNNOAVFVGIGVPTTDELMKDPTEKVFVYTK 274
Db      169 TSVVVPTQMAOHLQAGRIDGFCVGEFPWASAVOQDLFTWATSGAIWPDHKGVLGCTR 228
Qy      275 QMAEKYPTVYLAVTKALIRAAIWLADNNKRRKEAIEMLAQOQYGVADVEVLAASNGTFF 334
Db      229 EPEEQNPNTARALLMAVLEASRFIE-QSDHRRRSTQOLISGVYDIALDCTIEPRLLQY 287
Qy      335 EYEKDKRALPDNTFFPHG--ASYPSYSSAVWYLTQLRRGMINEFKPDNMYLDTAKNY 393
Db      288 SDGLGNQMDPHAVSFHDQGVNYPMLSDGMWFMQFRRWGLIRE-DDD--YLAVASRVQ 344
Qy      394 RPDYLAALKELVAEGKAAEDFPADTISIKPSQNFPTDKVPFDANKPNDYLAKEFAL 449
Db      345 QLDLYRQAAANL-----GIDAPSAT-LRSSQ--LIDGKXWDGSDPAGYARSFYL 390

```

```

RESULT 3
US-10-282-122A-66370
; Sequence 66370, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66370
;
; LENGTH: 399
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
; US-10-282-122A-66370

```

Query Match	9.8%;	Score 240;	DB 15;	Length 359;
Best Local Similarity	24.6%;	Pred. No. 4.9e-13;		
Matches 93; Conservative	68;	Mismatches 141;	Indels 76;	Gaps 14;

```

Oy 9 SKLLLTLLSASLAWNG-LTIAPDVGVANGLEKE-DLKFGFKITDPAFLAVALAKGFED 66
Db 9 SRDILIKLALLSAGALPLILSSLOQARAABEPAPRIGLPTTDTAPLLVAANGLFEA 68
Oy 67 EGLFVQLEAQA---NMKVVMDRVWANGELDGSHMLABAPLAASVGFTKADIEVPFSKGFN 123
Db 69 EG--IQAEPRVLLRSMAQVIEAFISQVNVVHILSMYTWAR--YGSKVPAKVANNHV 124
Oy 124 GNATIVSNEIWMQKPNILPEGGKPVNPIKADYLKEVEEKYKAEGKPFNNAMTPRAGSHN 183
Db 125 GSGLTVAPEI-----ADVRLQSGKSV-----AIPFWYSIHN 155
Oy 184 IKLRWMLAAGINPGVYSPRODISQIGA-----DALLSTVPBPQMPSTLEAGTIGYCV 238
Db 156 VVLQQLLRDNGI-----RAVSRAVGAALADEVNVLVLPSPDMPPALASRRINGYIV 207
Oy 239 GEPNQAQAVFKGIGVPTIDELIMKDTPEKVFSGTQMAEKYENTYLAATKALIRAIWTL 298
Db 208 AEPFALLAENLKVGVQRFPTGDVWRHNAACCVVRMHENHDERRFQMSQKVNALVKAQWM- 266
Oy 299 DADNNKNRKEALEMILAQ--KOYVAGDAVEVL-----AASMGTFEYEEKDKRA 343
Db 267 ---TREHRAEAQOLLSKAGANRRTYPHAPVGLGVLAPGAEEQQAUYLASGIRHADQERR 323
Oy 344 LPDENTFFRHGASYSBSYS 361
Db 324 I-IDQRY-----PYBSYT 335

```

```

RESULT 4
US-10-389-647-502
; Sequence 502, Application US/10389647
; Publication No. US2004003549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Canli
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UI2-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; PRIOR FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 502
; LENGTH: 399
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-502

```

Best Local Similarity 24.6%; Pred. No. 4.9e-13;  
Matches 93; Conservative 68; Mismatches 141; Indels 76; Gaps 14

```

OY 9 SKKLITLTSLSIAVMG-LTIADVGAVGLKEE-DLKEGFIKLTIDMAPIVAAGKEFED 66
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 9 SRBDILKATLALSAGALPLSLSSLOKRAAEEDAVRGYIPITDAPILVLVAHANGLEFEA 68
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 67 EGEFVOLTLEQA---NMKVMDVRVNGELDGSIMLAPALAAVSGFGTKADIEVPSMGFN 123
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 69 EG---IQABERVLLRSMAQYIEAFISGOYVNVHLSPMTVMAR--YGSKVPAKVVAHMHVG 124
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 124 GNAITYSNEIWHOMKENIPLBEGKRVHPIKADYLKPVVEKTKAAGKPPNMAMTPPASHN 187
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 125 GSGITVAPEI-----ADVRLQLGKSV-----AIPWYSIHN 155
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 184 IKLRVYLAAGSINPGVSPRODISOIGA-----DALSYTPPPQMPSTLEAGTIFGYCV 238
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 156 VVIQQLLRDNGL-----RAVSRAVGAALAADEVNLVVLVPSDMPALASKRIHGYIV 207
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 239 GEPNNOQAVFKIGIVITDEBLAMTDEKVGAVGVIKQMAEKYKPYTYLAVTKALIRAAIWL 298
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 208 AEFPNLALBMLKVGQRFTGQVWNHNAACVFFGHEHDLERRPQMSQVNVVAYIKADLM- 266
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 299 DADNNKKRKALIMLAQ---KQYVADAVEI-----AASNGTFFYEYKDKRA 343
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 267 ---TREHRAABAQLLSKAGANRYTPHAPVELGRVLAPGAERQOAVYLSAGAIRHADQERR 323
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

OY 344 LPDPNTFFRHHGASYSYS 361
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 324 I-DFQPY-----PYPSYT 335
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

```

RESULT 5
US-10-282-122A-49807
; Sequence 49807, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/267,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16

```

```

? Remaining prior identification data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 78614
? SOFTWARE: Patencntrn version 3.1
? SEQ ID NO: 49807
? LENGTH: 396
? TYPE: PRT
? ORGANISM: Burkholderia fungorum
? OS-10-282-122A-49807

```

Query Match	9.6%;	Score 234.5;	DB 15;	Length 396;
Best Local Similarity	23.3%;	Pred. No. 1.6e-12;		
Matches 105; Conservative	77;	Mismatches 150;	Indels 119;	Gaps 18;

```

QY      9 SKKLLTLLSSLSLWMLGTLPVGAVGKLEKED-----LKFGFKLTDMAPLAVAEKGF 64
Db      7 SREWMKLSMFLVAG--AAPLLAALNMRAANPNAPVRIGYLPITDAPLVAHNNGYF 64
QY      65 EDEGLFVLEA--QANKVMDRVNNGELDGSMTLAPAPLAASVGGTAKDIEVPSGFN 123
Db      65 DASGLAVEKPTLLRSWAOLVEAF.LSGQVVVHLLAPMTIMAR--YGSQAPAKVVAAMHVN 122
QY      124 GNAITVSNELWQMKRNIPLBEGGKPRPHIKADYLKPVVEKTKAEKGPRNMAFPGASHN 183
Db      123 GSAALTVAPDI-----GKLGELGGKV-----AVPFWYSIHN 153
QY      184 IKRWYLAAGINPGYSPRODISQIGADAL-LSVTEPPOMSESTLEGTIFYGVSEPW 242
Db      154 VVVVGHMLRAQGL-----VPLEKDEBLKANEVRLLVMSPSDMPALMSROJAGTVAEPF 208
QY      243 NQOAVFKGICGVITDEBLMKD-----TREKVFCTKQAEKPYNTYLAVTAKALIRAA 295
Db      209 NAAAEELKVGKYLRFPTGVWKNHACCVFMHERDLTERAASOK-----VDAAVYKAQ 261
QY      296 IMLDADNNKQRKAEITMLA---QKOYGAADVVLAASNNGTFEYEXKDKALDPNFFR 352
Db      262 VMTIRA---HPDEAOLLSKSGNNHTPISANVLT---VTLAPPGDGGRIALD----- 308
QY      353 HGASYPSSAVVYLLQLRRMGMINFEKPDNYLDTAKCV--YRPDIYLAARELV----- 406
Db      309 -----RAITHADWH-----AKRIDFGYPRPAYTEBELVRLTKA 341
QY      407 --AEGTAK-----AEDPADTSIKPS 425
Db      342 TQVEGNAOFLQDLPAFVARDLVDRFVYKS 372

```

```

RESULT 6
US-10-282-122A-44527
? Sequence 44527, Application US/10282122A
? Publication No. US20040029129A1
? GENERAL INFORMATION:
? APPLICANT: Wang, Liangsu
? APPLICANT: Zamudio, Carlos
? APPLICANT: Malone, Cheryl
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Karl
? APPLICANT: Zyskind, Judith
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John
? APPLICANT: Carr, Grant
? APPLICANT: Yamamoto, Robert
? APPLICANT: Forsyth, R.
? APPLICANT: Xu, H.
? TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
? FILE REFERENCE: ELITRA.034A
? CURRENT APPLICATION NUMBER: US/10/282,122A
? CURRENT FILING DATE: 2003-02-20
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26

```

```

PRIORITY APPLICATION NUMBER: 60/230,335
PRIORITY FILING DATE: 2000-09-06
PRIORITY APPLICATION NUMBER: 60/230,347
PRIORITY FILING DATE: 2000-09-09
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/267,636
PRIORITY FILING DATE: 2001-02-09
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 76614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 44527
LENGTH: 257
TYPE: PRT
ORGANISM: Acinetobacter baumannii
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (7)..(7)
OTHER INFORMATION: X=any amino acid
US-10-282-122A-44527

```

Query Match	9.3%	Score 227;	DB 15;	Length 257;
Best Local Similarity	23.5%;	Pred. No. 4e-12;		
Matches 69;	Conservative 52;	Mismatches 122;	Indels 50;	Gaps 77;

```

QY 9 LQSHSLAAPPLAAVGTGKAD-----IEVPSMGNGNAITVSNINQM-----KPN 141
Db 1 LPAHAKXSAMLPAAMG---ADQIGALQPLVLVSKRRAFISLSQKLIQQLAQESDNA 56
QY 142 PLEGGKPVHPIKADYLKEVEKYKAEGKPFNMAMTTPAGSHNITKLRYMLAAGGIPGYS 201
Db 57 QTTAQKVIQYIQDH-----TLSLAHVFKSHIHHCYKREMLALA-----95
QY 202 PQGDISGQIGAPALLSVTPPQMSSTLEAGIIPCYCGEPNQOAVFKGAVITDEEL 261
Db 96 ----DSRIAQTLKLKALPPPMVEALNNHVIDFCVGEPPNTOGELLGSLKIVCSSQDI 150
QY 262 WKDPEKYGVGTQWAEKXPNTYLAVTKALIRAAIMLDADNNKRNKEAIEMLAQKQYVA 321
Db 151 IPNVADKVLAVTQEWAEQHPTVLALTJAINKAQ--KELSLKXDFAPILTKL-----200
QY 322 DVEVLAASNNGIFFEYKDKKALPDENTFPHGSGSYSSSAWYLTQLRMRG 374
Db 201 -VEFGIVRHCSEVHVDKYMIQNIYCVLVKEMAAQPEDFHLFOQMKQ 252

```

RESULT 7  
US-09-738-626-4906  
Sequence 4906, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738, 626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16

```

; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4906
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4906

```

Query Match 9.3%; Score 226; DB 9; Length 294;

Best Local Similarity 23.6%; Pred. No. 6,1e-12; Matches 78; Conservative 56; Mismatches 152; Indels 44; Gaps 6;

```

QY 1 MKTIIRSSSKRLLLTSLASLAWGLTTPDVGAVGLEKED---LRFGIKLTDMAPLAY 57
DB 1 MTHILFDSRRFLQLGAFASLSTLAGAARYVTSTSNNEPADNPLTIGVPIGASAPIAI 60
QY 58 AAEKGFEDEGLVQLEAQAQNVMDRVVNGELDSHMLAPPLAASVGF-GTKADIEY 116
DB 61 ADALGFEKKGAVVTLTKYSGMSDLWTAYATEQLDVAHMLSPMTVAINGVTNASHPTET 120
QY 117 PFSGFGNNAITVSNELHMQMKNPILLEGKPVHPKADYLKRVKYEKGEPMAMT 176
DB 121 SFQNTNGAHTLASKHYGSVNSAADKQ-----MYLGIPFEYSV- 160
QY 177 FPAQSHNKLRYMLAAGINPGYSPPODISGOIGADALLSVTPPMQESTLEAGTIFGY 236
DB 161 ----HALLRDVLVNSAVDP-----IADLEKRLRPDMVAQGLVEGIDEG 202
QY 237 CGEPMNOQAVFGIGVPVITDEELMKDPEKVPFGVTKQAEKYEPTYLAVTALIRAI 296
DB 203 IGPFPNEBAISNGSSRIMWLTQKQWDKPCAVAAKEKAEHPAQAQVNLALBEASA 262
QY 297 WLDADNNKREKALEMLAOKQYGVADVEYL 326
DB 263 IL--SNPAQFDSARTLSDEKYLNPATL 290

```

## RESULT 8

```

US-10-282-122A-68363
; Sequence 68363, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578

```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68363
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-68363

```

Query Match 9.1%; Score 222.5; DB 15; Length 372;

Best Local Similarity 23.1%; Pred. No. 1,8e-11; Matches 90; Conservative 71; Mismatches 139; Indels 89; Gaps 17;

```

QY 42 LRFGIKLTDMAPLAYAAEKGFEDEGLVQLEA--QANKVMDRVVNGELDSHMLAPA 100
DB 17 VAGIVLPIDATPPLVAHNNGLFEAGIKAEKRPVLLRSQAQVIEAFISQVAVIHLSFM 76
QY 101 PLAASVGFCTKADIEVPSFGNGNAITVSNELHMQMKNPILLEGKPVHPKADYLKEV 160
DB 77 TWAAR--YGSKPAKAVANNHVGSGSLTVAAPDI-----SAVKQLGGKTV----- 118
QY 161 VEKYAEKGEPMAMTFPAGSHNKLRYMLAAGINPGYSPPODISGOIGADAL--LSVT 219
DB 119 -----ALPFWYSIHNVVYQOLLNDNGLP--VSKP--ANAOLANEVNVLVL 161
QY 220 PPMQESTLEAGTIFGYCGEPMNOQAVFGIGVPVITDEELMKD-----TPKEVFGV 272
DB 162 PPSDMPPLASGRVIAEFPNLAELKXGRVRFQGDVWRHAFACCVFMHEHDANN 221
QY 273 TKQMAEKYPNTYLAVTALIRAIWLDADNNKREKALEMLAQ--KQYGVADVEYLAAS 329
DB 222 RPEWSQK-----VYNAIVKAQW---TRDHRTEAALLSRAGPNKYPHEPAVLTKV 270
QY 330 M-----NGTFEKKDKRALPDNTFFPHGASYPSSAVWVLTQLRKGMIN 377
DB 271 LAPAAEDRAGYASGIRIQWDEKRI--DFQY-----PFSYTEB--LVRLKTTLL- 320
QY 378 EFKPDWVYLDTAQNYRPPDIYLAALKELY 406
DB 321 --EGDNTFLSGLDPAV-----AARDLV 340

```

## RESULT 9

```

US-10-282-122A-49957
; Sequence 49957, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

```



```

RESULT 12
US-10-282-122A-72603
; Sequence 72603, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangou

```

Qy 1 LILLTASALVAVGLTAPDVAVGKLEKEDKLFKFKIKLTDK--RPLVAAEKKEFEDEGL 69

Dp 15 ILTITISVLAVCSKSTSTSKNYNAASEKDSYTLKISENSDLCGAPQOALIEKGFDDVGL 74

Qy 70 FVQLEAQAMKV-----MDRVNGLDGHMLAPA---PLAAVGFGTKADIEVPE 118

Dp 75 -----KTKYVKIGQDTSNDLALNAKIKLDAINSMASTIQLA----- 111

Qy 119 SMGFGNALTITSNEIWHQKNPIPLEGGKPVHPIKADYLKVEKRYAEKGPFFMAATP 178

Dp 112 -----NGAKLKITGL- HTGCLQILFKNGK-----IKSAABLKGKKIGVTAV 152

Qy 179 AGSHNIKLRMYLAAAGINPGYSPPODISGQIGADALISTVPPQMSPTLEAGTIFGCV 238

Dp 153 AGSPALFAKRVVLAKSGLK-----VSDEKG-DVSEFVTYQSDQLQGVLDKGEVDAL 202

Qy 239 GEPMNQOAFPKIGVPVITDEBELMKDPTREK---VPGYTKQAEKPYTYLALVRLRA 295

Dp 203 GDE-DTEVLKQGYGFKTLANSTDGFKNEXCYAAVYSNDI VKKHVPAAKYTLAQA 261

Qy 296 IWLDAQNNGKREAIEMLAOKOYGVADVEVLAAASNG--TFEYEKDGRALPDFTNPRHG 354

Dp 262 NMV-----QKHKEEYDIDQLNQYVAGSKDSNLSLSTYTK----- 298

Qy 355 ASYPSYSSAV---WYLTQLRRMGMIN 377

Db 299 ---PSYSGAYDSFDTVASDLRKIGILS 322

## RESULT 13

US-10-282-122A-47676  
; Sequence 47676, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Foreysch, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIOR FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See file wrapper or PAM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 47676  
; LENGTH: 192  
; TYPE: PR1  
; ORGANISM: Burkholderia cepacia  
US-10-282-122A-47676

Query Match 5.7%; Score 140; DB 15; Length 192;  
Best Local Similarity 26.8%; Pred. No. 0.00027;  
Matches 42; Conservative 32; Mismatches 51; Indels 32; Gaps 4;

QY 42 LKGGFKLTDMALVAAEKGFFEDGGLV-QLAEQAQNKVMDRVNGLDGSHTLA 100  
DB 37 VRIIGYLPITDAALLVAHNNGYFASGELTYEQPKLRSMQGLVEAFLSQVNVVHLSPM 96  
QY 101 PLAAAGFGGTAKADIEVPFSGFNAGATVSNETIWHQMKPIPLEGGKPVHPRIKADYLV 160  
DB 97 TLMAR--YSGRAAKVYVAMNHVNGSLTYAPDI-----DSLRLGSKTY----- 138  
QY 161 VEKYKAGKPFNNAMTFPAGSHNIKLRVWLAAGINP 197  
DB 139 -----ALPFWYSIHVVVLQDMLREGGLVP 162

RESULT 14  
US-10-156-761-8511

; Sequence 8511, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:

; APPLICANT: OKURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 8511  
; LENGTH: 883  
; TYPE: PR1  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8511

Query Match 4.8%; Score 116.5; DB 14; Length 883;  
Best Local Similarity 21.3%; Pred. No. 0.41;  
Matches 114; Conservative 62; Mismatches 147; Indels 213; Gaps 35;

QY 6 RSSSKLLITVLSASLAVNGLTAPDVGA-----VGKLEKEDLKFGFKLTDMAP---L 55  
DB 15 RSLGRSLALTMTAVVGAAGLALSPASADSGPTTADPTFHAIGKGEV--YTQASGAFDF 72  
QY 56 AVAAKGGFFEDG- FVQLEQAQNKVMDRVNGLDGSHTLAAPLAASGFGTKADI 114  
DB 73 AELKATGF--DPGLDFDSLSR-----LSAATGRADVSV 105  
QY 115 E-----VP-----FSM-GFNGNATVSNET-----W--HQMKNIPLEGKPVHP 151  
DB 106 RWTGEVVEKSGPTTFSTGDNQFRLWDKLTIDHWVDMDHQEQTSQPVDTLQAGR-AVD 164  
QY 152 IKADYLVKPVVEKYKAGKPFNNAMTFPAGSHNIKLRVWLAAGINPGYSPQ----- 204  
DB 165 IKEYFEHF-----GGSNTHLR--WTPPGSKS---AVQOSAFRLPD 201  
QY 205 --DISGQIGADAL-----LSVPPQMP--STLQAGTIFGVC 237  
DB 202 GFDVNGALSAIVLGDGRTVRLDFQALASAPATVLDHLTVSGKPLSSVKT----- 254  
QY 238 VGEPMNQOAVFKGIGVPIYTD-----ELMKD-----TPEKVFQVTRQWARK 279  
DB 255 --DPSDARALLVGLGEPVVGKAGKAGVAVTYDGKALADTSGTPYKAF-----WSSG 307  
QY 280 YFN--TYLAVTKALIRAILWLDADNNKRNKEAIEMLAQOYVADDEVILA-ASMXGTBEY 336  
DB 308 -PNKSTYELRTE-----MADQVGRN-----ALPEYPRPOLTRKAMONLNGTWQF 351  
QY 337 -----EKDDKRALPDNTEFFRHGASVP-----SYSSAVVYVLTOLRWGMINERK 380  
DB 332 AGAKAGKEEPPGRTLA-----RIIVPYFVESQSGIERHEDRMV-----RTFTV----- 398  
QY 381 PDNWTYLDTRAKN-----YRPDIYLAAXELVNEGKAKAEDFPADT--SITKPSQ 426  
DB 399 PADMKVGVGKRLRLNFGAVNDWQSEVYVNGTK--VAEHKGVYDKFSDADVTDLAKPER 452

RESULT 15  
US-10-450-763-36085  
; Sequence 36085, Application US/10450763  
; Publication No. US20050196754A1  
; GENERAL INFORMATION:  
; APPLICANT: Hybee, Inc  
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES



```
FILE REFERENCE: 790CIP3/US
CURRENT APPLICATION NUMBER: US/10/450,763
CURRENT FILING DATE: 2003-06-11
PRIOR APPLICATION NUMBER: PCT/US01/08631
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: 09/540,217
PRIOR FILING DATE: 2000-03-31
PRIOR APPLICATION NUMBER: 09/649,167
PRIOR FILING DATE: 2000-08-23
NUMBER OF SEQ ID NOS: 60736
SOFTWARE: Custom
SEQ ID NO: 36085
LENGTH: 1355
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (287)..(307)
OTHER INFORMATION: Elongation factor 1 beta/beta'/delta chain proteins domain
OTHER INFORMATION: Identified by eMatrix, accession number BL008248, p-value=8.419e-
US-10-450-763-36085
```

Query Match 4.6%; Score 113; DB 18; Length 1355;

Best Local Similarity 21.6%; Pred. No.1.7; 196; Indels 152; Gaps 25;

Matches 110; Conservative 51; Mismatches 152; Gaps 25;

```
QY 2 KTIIRSSSKLLITLISA-----SLAVWGLTIAPDVGAVKL--EKEDLKEFTK 48
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
790 RTDVAESRKGLTRSAQETVKHSDLFSSSSPMWKGTKPRTKTVLSLPDEED-----K 843
QY 49 LIDMAPLVAAEK-----GFPEDEGLF-----VLEAQANMKV 81
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
844 MEDSKPLSQAPQKEVGKCDPDAPKSTGVFDDELLFSHKLQKNDPPVDLFAGTKTK 903
QY 82 VMDRVVN--GELDSHMLAPA--PLASVGFGTADIEV--PFSMGFNQNAITVSEIM 134
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 LLEPIVGSIFGDEDDDLFSSAKSQPLVQEKRAVKKDSVDSFKQKHPESTIQSKKKG 963
QY 135 HOMKPNIPLEGKRPVPIKADYLVKPYEKYKAEKPPNNAMTFPPAGSHNIKLRWYLAAG 194
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
964 HMMKEPTP-QANLAINPAA--LLPFAAQISEVKVLPDLAPSEH--RRSHGLESV 1017
QY 195 INPGYSPQDISGQIG-----ADALISTTP-----POMESTLEAGTIFGYCVGB 240
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1018 VLPG-----SGEAGVSFDLPQAADTLHSANKSRVKRGKRPPQTRAAARRLAA----- 1064
QY 241 PNNQOAVFKIGVPIVITDELMWDTPKEVFGVTKQNAEK--YPNITYLAVTKALIRAIWL 298
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1065 --QESSEADMSIP-----RPIAQADGALISPNCH---RPQLRA----- 1099
QY 299 DADNNKRRKAEIEMLAQK---QYVGADEVLAASNGTFEYKODKRALPDNTEFFRHG 354
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1100 -ASGEDSTEELAAALAAAPWEGGVPEVDVSPFAKSL-GHSRGAD---LFDGDIFFSTG 1153
QY 355 ASYPSYSSAVVYLTQLRKMGMINFEKPDWYLDTANVYRPDIYLAALAEVAEGKAE 414
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 TGSQS-----VERTKPK--AKIENPANPV-----GKAKASP 1184
QY 415 DFPADTSIKPSONFFIDKVPFDANKPNDY 443
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1185 MFPALGEASSDDLDFOSAKPKPAKKTNP 1213
```

Search completed: September 16, 2005, 01:02:42

Job time : 178 secs

**This Page Blank (uspto)**